Multistakeholder Recommendation

Part 2

Objectives

- You should come away from this part of the session
  - Understanding the relationship between fairness-aware recommendation (FAR) and multistakeholder recommendation
  - Understanding different definitions of fairness and the contested nature of the term
  - Understanding the difference between fairness in classification systems and in recommendation
  - Understanding some algorithms for implementation and evaluating FAR systems

Acknowledgements

- Some material is drawn from the upcoming RecSys tutorial
  - Fairness & Discrimination in Recommendation & Retrieval
    - Burke, Diaz, Ekstrand
- Also, much work by my students Nasim Sonboli and Himan Abdollahpouri
  - That Recommender Systems Lab, CU Boulder

Fairness

WHY CAN'T I STAY UP LATE? YOU GUYS CAN!
IT'S NOT FAIR!
THE WORLD ISN'T FAIR, CALVIN.
I KNOW, BUT WHY ISN'T IT EVER UNFAIR IN MY FAVOR?
WATTERSON 8/1

Fairness = S_t

- Many users and providers may not (probably don't) care about fairness
  - "I want what's best for me." = Calvin's "Unfair in my favor"
- Fairness is a system concern
  - May be a matter of legal requirement
  - May be a matter of organizational mission
  - May be a matter of user / provider retention
    - essential in a multisided platform

Discrimination / inequality can be OK

- Charge students $274 registration
  - typically younger
- Charge professors $681 registration
  - typically older
- No senior discount!
- Age discrimination!
- But we're OK with that

What counts as fair/unfair?

- Unfairness = an unjustified harm / benefit
- Harm / benefit
  - What kinds of harms / benefits are associated with a recommender system?
  - Not always obvious ones
- Justification
  - What kinds of unequal harms / benefits can be justified?
  - On what grounds?
  - Note the word "justice" is lurking here
- These are not computer science questions

Where do we look for answers?

- Fairness is a social concept and inherently normative
  - Selbst et al. 2019: fairness "can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms"
- Engaging with these problems requires engaging with many disciplines:
  - Law
  - Ethics / philosophy
  - Economics
  - Sociology
  - Political science
  - Feminist / post-colonial studies
  - Etc., etc.

Considerations

- What category of harm are we interested in?
- What is the justice construct we are going to use relative to that harm?
- To whom are we applying the construct?
- What specific type of outcome will we examine?
- What will be our metric on that outcome?
- How will we optimize to improve that metric in our recommendation results?

Distributional harms

- Occurs when someone is denied a resource or benefit
- In recommendation
  - a female student doesn't get a recommendation for a computer science class
    - and a similar male student does
  - a new seller's items are not recommended to potential buyers
    - and more established sellers' items are
  - etc.

Consumer side (C-fairness)

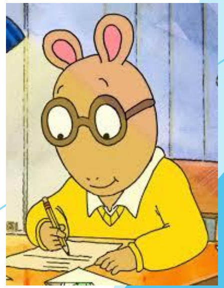
- Site may wish to be fair to the consumers of recommendations
  - Job seekers
- Example: male job seekers should not get better / different recommendations than female
  - Might be a legal requirement

Provider fairness (P-fairness)

- Fairness relative to items being recommended
- A book seller might care about being fair to authors
- Do minority authors have a fair chance of being recommended?

Fairness across items

Because of creators / owners



Representational harms

- Distributional not the only kind of harm in recommendation
- Representational harms arise when someone is represented incorrectly in the system or to its users.
  - Misgendering
  - Racial miscategorization
  - Stereotyping (esp. reinforcing negative stereotypes)
  - ‘Inverse’ representational harms: who shows up when searching for ‘ceo’?
- Appearance of items in recommendation lists
  - is a form of representation
  - Noble's "Algorithms of Oppression" discusses representational harms esp. in search engines

Phenomenological harm

- Harm can occur when data subjects perceive themselves as powerless
  - and thus vulnerable
  - when personalized systems "know too much"
- This is the "uncanny valley" of recommendation and personalization
- May be particularly experienced by groups who are disempowered in other ways
  - separate from privacy concerns
  - Burke & Burke, "Powerlessness and Personalization". IJAP forthcoming.

What questions so far?

Exercise

- Continuing with our out-of-school time activities application
- What kinds of disparate harm / benefit might happen in this system:
  - To consumers?
  - To providers?

What's justified?

- Welfare economics gives us four types of justification for fairness
  - Moulin H., Fair division and collective welfare, MIT Press, 2004
  - Fairness required by **exogenous right**
    - Legal requirements or standards
  - Fair **reward**
    - Fair rewards
    - Like a bonus
  - Fair **compensation**
    - Affirmative action
  - **Fitness**
    - Best match between resource and recipient

Fairness as exogenous right

- External dictate about what each party is entitled to
  - Usually legislation
  - You can be sued if someone can prove you were unfair
- Usually the standard is rough equality between protected group and others
  - various standards of proof and impact within the legal system
- This is the usual case in discussions of machine learning fairness
  - but not the whole story

Fairness as Fitness

- An outcome might be fair if it allocates resources to those best able to use them
- Online multivendor site with vendors A and B
  - A sells mass market electronics
  - B sells pricey audiophile gear
- Most customers buy from A
  - Small number of aficionados buy from B

Fairness as Fitness, cont'd

- In recommending products to customers
  - Is it fair that A's and B's product appear with the same frequency for all users?
  - No, because there is a fitness consideration
- The typical user is not a good customer for B
  - A fair distribution of recommendations across users
  - Takes the fitness of the customer into account
  - B should get the right kind of customer
    - Even if they are fewer

Fairness as Reward

- An outcome might be fair if it allocates resources as a reward to contributions made
- Online multivendor site with vendors C and D
  - C is a cut-price brand that doesn't do much marketing
  - D is a well-known brand that does a lot of marketing
- Customers are attracted to the site by D
  - But sometimes buy from C when there's a good deal
- Might make sense to give D a bigger share of the recommendations
  - Reward for bringing in business

Compensation

- An outcome might be fair if it compensates a party for costs, losses or risks
- In this setting, a member of the protected group would be expected to get greater utility from the system
  - Than an unprotected group member
- Affirmative action is a well-known example in the US context
  - Deliberate inclusive action in hiring, promotion, etc. as compensation for historical lack of opportunity

What questions so far?

Exercise

- Consider our recommender system for out of school activities
  - and the types of harm you discussed earlier
- What kind of fairness justification is called for in each case?
  - Exogenous right
  - Reward
  - Compensation
  - Fitness

Protected group / class

- Protected attribute
  - Gender, religion, race, sexual orientation, etc.
- Defines a protected class
  - Usually but not always a minority class
- Goal
  - Decisions should be independent of the protected attribute
  - Protected and unprotected cases treated the same if that’s the only difference
- Oversimplification of real world complexities
  - need to start somewhere!

Measuring Fairness

- Individual fairness says similar individuals should be treated similarly
  - Two applicants with the same ability to repay a loan should receive the same decision
  - Harm that can't be justified on an individual level
- Group fairness says each salient group of people should be treated comparably.
  - Black loan applicants should not be denied more often than white
  - Often concerned with a protected class or sensitive characteristic
    - In U.S. context, anti-discrimination law provides this
  - Harm that can't be justified on a group basis

Why is individual fairness insufficient?

- Fundamental reasons: historical discrimination + measurement impossible
  - Measures of individual merit are skewed
- Prospective outcomes may vary for social reasons
  - Example: Academic standardized tests predict socioeconomic status.
  - Scores conflate aptitude and preparation
- The idea of two profiles being "the same except"
  - is very problematic for groups that have experienced discrimination

Group Fairness Concepts

- **Disparate treatment**: members of different groups are treated differently
  - Applying different standards to people of different ethnicities
- **Disparate impact**: different groups obtain different outcomes
  - Example: Men pass the employment test for firefighters at a higher rate than other genders
  - Foundation of much U.S. anti-discrimination law
- **Disparate mistreatment**: different groups have different error rates
  - A risk assessment tool is more likely to misclassify a black defendant as high-risk

Important results

- Attempts to remove disparate treatment / disparate impact / disparate mistreatment
- Will generally be incompatible with each other
  - Z Lipton, J McAuley, A Chouldechova. Does mitigating ML's impact disparity require treatment disparity? 2018
- You have to decide what's important in your application
- There are a lot of different things you can measure

Consumer-side outcomes

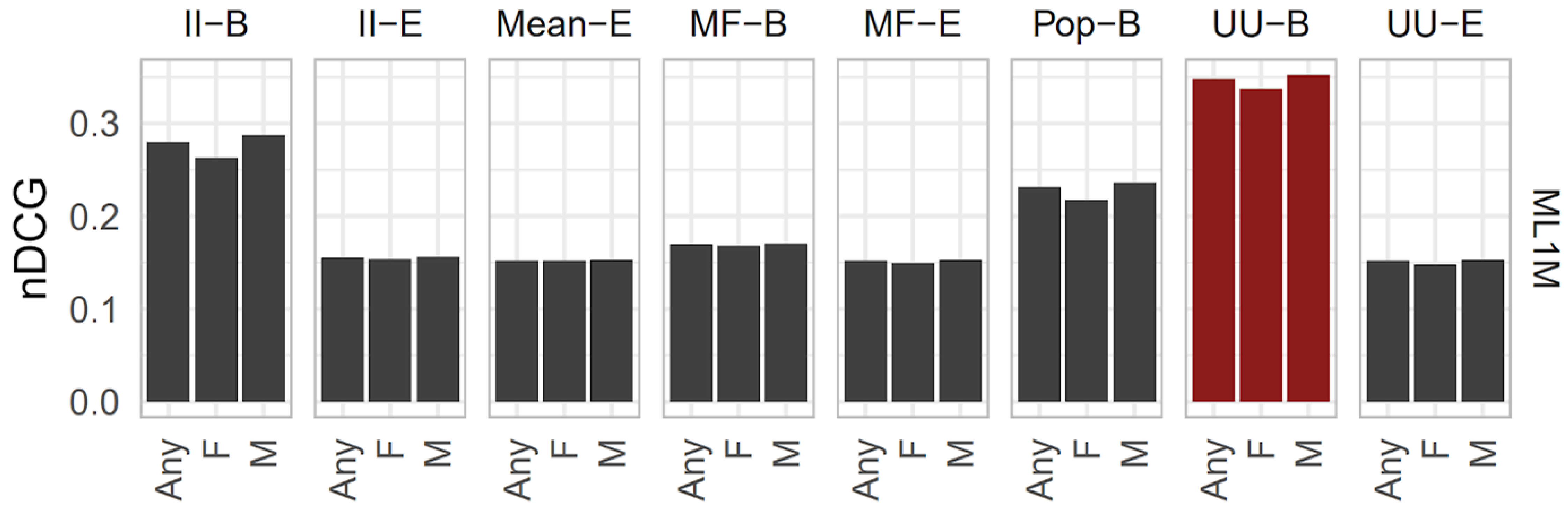
- What is recommended?
  - **Do protected and unprotected groups get "good stuff" recommended to them?**
  - not much research in this area
  - Some lawsuits
    - ACLU vs Facebook on job ads
  - In general, need some idea of where the harm lies
    - premise of personalization
- Quality of recommendations
  - **Do protected and unprotected groups get similar quality of results?**
  - Ekstrand, et al. All the Cool Kids, Where Do They Fit In? FAT* 2018
  - Yao & Huang Beyond Parity NeuroIPS 2018

Recommendation quality

- Do different groups experience different error rates?
- Why?
  - Group sizes / distributions
  - Niche tastes
  - Algorithmic properties
    - inductive bias
  - Item distribution
- Long-tail properties of recommendation data sets
  - may explain some aspects

Differential quality

Different nDCG across demographic groups



Ekstrand, M et al. All The Cool Kids, How Do They Fit In: Popularity and Demographic Biases in Recommender Evaluation and Effectiveness. FAT* 2018

Provider-side outcomes

- Fairness of exposure
  - Are providers / items getting fair exposure?
  - Any of the exposure metrics could be relevant
- Fairness of audience
  - User / market diversity
    - not well studied
- Fairness of accuracy
  - How accurate are predictions of provider's items compared with others?
    - not well studied
    - Example: Ekstrand, M. et al. Exploring author gender in book rating and recommendation. RecSys 2018

Metrics

- A wide variety of metrics are possible
  - because of the different outcomes of interest
- And different ways of comparing outcomes
- Examples
  - group accuracy via nDCG compared by difference (Ekstrand et al. 2018)
  - exposure via protected group precision compared by ratio (Burke et al., 2018)
  - probability of item recommendation compared by KL divergence (Yang & Stoyanovich, 2017)
- Way too many to list

Metrics: what to do?

- Ideally, we would work from
  - the harm,
  - the outcome that reflects that harm, and
  - the relevant justice construct
- All validated based on real-world experience
  - Doesn't usually happen
- Often metrics are derived in a vacuum
  - or at least without these matters being made explicit
  - (Full disclosure) I do this, too

What questions so far?

Exercise

- Consider our recommender system for out of school activities
  - and the types of harm you discussed earlier
- What kinds of outcomes would you look at in each case?
  - content
  - accuracy
  - exposure
  - audience

Algorithms

- Fairness-aware recommendation is a type of multistakeholder recommendation
- Similar algorithmic approaches
  - Multi-criteria optimization
  - Regularization
  - Re-ranking

Regularization

- Typical matrix factorization methods have an objective incorporating regularization to control overfitting
  - limiting the "size" of the factors

$$\text{Minimize } J = \frac{1}{2}\sum_{(i,j)\in S} e_{ij}^2 + \frac{\lambda}{2}\sum_{i=1}^{m}\sum_{s=1}^{k} u_{is}^2 + \frac{\lambda}{2}\sum_{j=1}^{n}\sum_{s=1}^{k} v_{js}^2$$

- We can use regularization to enforce other types of constraints including fairness-related ones
  - Much work along these lines
  - Kamishima
  - Yao and Huang

Balanced Neighborhood SLIM

- SLIM = Generalization of nearest neighbor
- Instead of discrete neighborhoods
  - We predict based on personalized regression equations
  - The coefficients define “near” and “far” items
- Balanced neighborhod constraint
  - says that weights of protected and unprotected groups should be similar
  - use this as an additional regularizer
    - can be used for both C-fairness and P-fairness

$$\hat{s}_{ij} = \sum_{k \in U} w_{ik} r_{kj},$$

$$\min_W \frac{1}{2} \|R - WR\|^2 + \lambda_1 \|W\|^1 + \frac{\lambda_2}{2} \|W\|^2,$$

$$L = \frac{1}{2} \|R - WR\|^2 + \lambda_1 \|W\|^1 + \frac{\lambda_2}{2} \|W\|^2 + \frac{\lambda_3}{2} \sum_{i \in U} \left(\sum_{k \in U} p_i w_{ik} \right)^2$$

Regularization on W

Regularization on neighborhood balance

Fairness is equal probabilistic exposure

- Results are fair if P(R | a) = P(R | â)
  - where a is the item with sensitive feature and â is the item without
- To achieve
  - regularization term penalizing non-independence
  - Kamishima, et al. "Recommendation Independence" FAT* 2018.
- Note
  - Form of "counterfactual fairness"
  - â might not exist or even be possible

Fairness is list-proportional ranking

- Meaning that the protected class should be equally prevalent in all top-k prefixes of the recommendation list
  - Yang & Stoyanovich. Measuring fairness in ranked outputs. ICSSDM 2017.
- Try to achieve through re-ranking

$$\delta_i = \frac{1}{\log_2 i}$$

$$\mathrm{rND}(\pi) = \frac{1}{\mathcal{Z}} \sum_{i=1}^{|\mathcal{D}|} \delta_i \left| P(a|\pi_{\leq i}) - P(a|\mathcal{D}) \right|$$

$$\mathrm{rKL}(\pi) = \frac{1}{\mathcal{Z}} \sum_{i=1}^{|\mathcal{D}|} \delta_i D_{\mathrm{KL}}(P(A|\pi_{\leq i}) || P(A|\mathcal{D}))$$

$$\mathrm{rRD}(\pi) = \frac{1}{\mathcal{Z}} \sum_{i=1}^{|\mathcal{D}|} \delta_i \left| \frac{P(a|\pi_{\leq i})}{P(\overline{a}|\pi_{\leq i})} - \frac{P(a|\mathcal{D})}{P(\overline{a}|\mathcal{D})} \right|$$

Other re-ranking approaches

- Pairwise fairness
  - P Sapiezynski, et al. Quantifying the Impact of User Attention on Fair Group Representation in Ranked Lists. WWW 2019.
- Rank-aware calibration
  - Beutel, et al. Fairness in Recommendation Ranking through Pairwise Comparisons. arXiv 2019.
- Amortization over time
  - Biega, et al. Equity of attention: Amortizing individual fairness in rankings. SIGIR 2018.

Conclusion

- Fairness-aware recommendation
  - subclass of multistakeholder recommendation
  - in which there is some type of system fairness goal
- Diverse problem space
  - Variety of possible parties: Consumers / providers / possibly others
  - Variety of possible harms / benefits; Usually distributional
  - Variety of justice constructs: not always equality by exogenous right

Open Problems

- Similar to multistakeholder problems generally
- Transparency
  - How to explain fairness-aware recommendations?
  - Interesting HCI problem
- Tradeoffs
  - Especially since there are multiple possible fairness metrics
    - even for a given justice construct
- UX
  - Can stakeholders audit to see whether fairness is achieved?
- Dynamics
  - Understanding positive feedback loops and temporal balancing of outcomes
- Intersectionality / Subgroup fairness
  - Not just one protected feature / protected group
  - Particular intersections of identity have their own fairness concerns

References

- Beutel, et al. Fairness in Recommendation Ranking through Pairwise Comparisons. arXiv 2019.
- Biega, et al. Equity of attention: Amortizing individual fairness in rankings. SIGIR 2018.
- Burke, V. & R. Burke. "Powerlessness and Personalization". International Journal of Applied Philosophy, forthcoming.
- Ekstrand, M. et al. "Exploring author gender in book rating and recommendation". RecSys 2018
- Kamishima, T. et al. "Recommendation Independence", FAT* 2018
- Lipton, Z. et al. "Does mitigating ML's impact disparity require treatment disparity?" NeuroIPS 2018
- Moulin, H. *Fair division and collective welfare*, MIT Press, 2004
- Noble, S. *Algorithms of Oppression: How search engines reinforce racism*. NYU Press, 2018
- Sapiezynski, et al. Quantifying the Impact of User Attention on Fair Group Representation in Ranked Lists. WWW 2019.
- Selbst, et al. "Fairness and abstraction in sociotechnical systems." FAT* 2019.
- Yang, K. & J. Stoyanovich. "Measuring fairness in ranked outputs". ICSSDM 2017.

- See also references for fairness tutorial
  - https://fair-ia.ekstrandom.net

Acknowledgements

- Michael Ekstrand, Fernando Diaz (Fair IA tutorial team)
- Nasim Sonboli, Himan Abdollahpouri (That Recommender Systems Lab, CU Boulder)
- Bamshad Mobasher (and the other members of the Web Intelligence Lab, DePaul University)

Thanks!