Fair and Efficient Social Decision-Making

CSCI 699

Embedded EthiCS Module

Evi Micha

Credit for the slides : Nisarg Shah

Embedded EthiCS

- Ethical reasoning skill a must for computer scientists

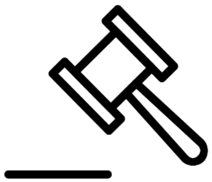
- Embedded EthiCS™
  - Distributed pedagogy approach initiated at Harvard CS
  - Embedding ethical thinking and reasoning into CS courses

- Goals
  - To show CS students the extent to which ethical issues may arise when designing and deploying algorithms
  - To familiarize students with approaches to ethical design
  - To allow them to practice reasoning about ethics, articulating their positions, and incorporating their ideas into the systems they design

Algorithms Making Decisions

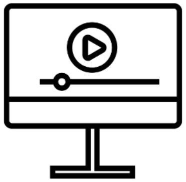
Loans

Bails



Self-Driving Cars

Ads



Hiring

Organ Exchange

Sources of Unfairness/Bias

- Bias in training/input data
  - Historical bias
  - Representation bias
  - Measurement bias
  - Simpson's paradox
  - …

- Bias in the algorithm
  - Direct discrimination
  - Indirect discrimination
  - Statistical discrimination
  - Justified vs unjustified discrimination
  - …

Types of Unfairness/Bias

- Outcome Fairness
  - Fairness in the outcomes produced by the algorithm
- Procedural fairness
  - Fairness of the algorithmic procedure
- In this course, we mainly focused on outcome fairness
  - We assumed that an agent's utility in a specific instance depends only on the outcome produced in *that* instance
  - But more generally, the utility may depend on the algorithm itself
  - Example: when I vote for candidate A and they lose, I may be unhappy, but may be more accepting of the outcome if I know that a fair rule like plurality was used to select the winner

Definitions to Fairness

- Individual fairness
  - Individuals are treated fairly
- Group fairness (stronger than individual fairness)
  - Groups of individuals are treated fairly
- Group fairness (weaker than individual fairness)
  - *On average*, groups are treated fairly (but individuals members in those groups may be worse off)
- Extensions
  - Different entitlements, history, demographics, legal constraints, …

Economic Approaches

- Individual fairness
  - Proportionality: each individual gets their fair share
  - Envy-freeness: no individual envies another individual

- Group fairness
  - Core: each group of individuals gets their fair share
  - Group envy-freeness: no group envies another group
  - Stronger than individual fairness
  - There are also similar group fairness notions that are weaker than individual fairness

ML Approaches

- Popular fairness definitions
  - Demographic parity
  - Equal opportunity
  - Equalized odds
  - Calibration
  - Typically, pre-defined groups and binary outcomes

Breakout Activity 1:
What does fairness entail?

Breakout Activity 1

- *What does fairness entail?*
  - You'll be divided into breakout groups
  - Each group will receive a hypothetical scenario, in which they will be tasked with making a decision that affects several entities
  - Each entity can be described with various features
    - E.g. a person can be described using their race, gender, education history, marital history, credit score, whether they're afraid of heights, …
  - Most features would be *irrelevant* for the decision at hand

Breakout Activity 1

- Goals
  1. Identify the features which are relevant for the decision at hand
  2. Partition these features into two classes:
     - **Should Use:** For a good decision, one should take these features into account
     - **Must Avoid:** For fairness, the decision must not discriminate based on these features, as much as possible

- For example, a bank deciding whether to accept a loan application from an individual may consider "the number of previous loans defaulted" under *should use,* but race or gender under *must avoid*

Scenario 1

We are in the middle of a pandemic (knock on wood!). Thankfully, vaccines have been discovered, and countries have begun inoculating their citizens.

You are part of a global organization that has secured a large pool of vaccine doses. Your goal is to decide how to divide these doses between the various countries that have approached you. Assume you are giving the doses free of charge.

1) Spend a couple of minutes discussing what your objectives should be, e.g. , fairness, maximizing the number of lives saved, minimizing the number of hospitalizations, etc.

2) Then list the features of the countries that you believe your final decision should or must not depend on.

Example features to get you started (please come up with more): Size of the population, the age distribution, gender ratio, GDP, current rate of vaccination, average willingness of the population to get the vaccine…

Scenario 2

You are the hiring manager at Fair Bank.

You have posted the job and have received applications from excellent candidates.

Your goals (feel free to consider other goals) are to hire a candidate who:

-is well-mannered and friendly (this is a customer - facing rule);

-has a basic understanding of finance and economics;

-is reliable in terms of showing up to work on time every day; and

-is efficient, hardworking, and dedicated

After discussing with your group for a few minutes, list the features of the candidate that you believe your final decision should or must not depend on.

Example features to get you started (please come up with more): Age, race, gender, neighborhood of residence, country of residence, education history, number of kids, marital status (e.g. current marital status or number of divorces)…

Synthesizing Thoughts from Activity 1: What does fairness entail?

Participatory Budgeting

Participatory Budgeting

- Setting
  - Infrastructure projects proposed across a city
    - Each project p has a cost c_p
  - Budget B reserved by the city for funding these projects
    - A subset of projects S can be funded if $\sum_{p \in S} c_p \leq B$
  - Residents vote over the proposed projects
    - E.g. they could be asked to…
      - Select the top 3 projects they like (3-approval)
      - Rank the projects by how much they like them (ranking)
      - Rank the projects by "value-for-money" (VFM)
      - Select the best subset of projects according to them which fits the budget B (knapsack)

Goals

- Many goals not related to the final decision-making
  - Ensuring participation by diverse communities
  - Facilitating community discussion for filtering projects and to ensure an informed vote later on
  - …
- Final decision-making should balance the allocation of funds between…
  - Preferences of different sub-communities
  - Geographical regions
  - Category of projects (education, healthcare, parks, roads, …)
  - Low-cost versus high-cost projects
  - …

Approaches to PB

- Welfare maximization
  - Elicit or estimate the happiness of the community from each project
  - Select a feasible subset of projects maximizing the total happiness
  - For example, if each resident votes for their top 3 projects, select a feasible subset of projects to maximize the total number of votes

- Fairness: the core
  - Out of all residents N, there should be no $S \subseteq N$ such that by using their proportional share of the budget $B \cdot |S|/|N|$, they could fund a subset of projects which would make each of them happier than under the current decision

Breakout Activity 2:
How should the
public budget be allocated?

Breakout Activity 2

You are a city official in Utopia City. Your city has conducted a participatory budgeting election, the residents have voted , and you are in charge of making the final decision.

Budget: $700,000

(A) Kitchen Renovation at Primary School ($400,000)

(B) Installing lights in playgroups for safety play ($300,000)

(C) Building a new public park ($600,000)

(D) Dedicated bike lanes in a neighb or hood ($300,000)

Votes: During voting, residents were shown the four projects but not their costs. They were asked to rate each project on a scale of 1 (least liked) to 10 (most liked) Here are 7 responses, for simplicity:

	A	B	C	D
R1	10	9	8	7
R2	1	3	10	2
R3	9	7	6	5
R4	3	2	8	4
R5	10	8	9	6
R6	9	6	3	4
R7	1	5	9	2

Synthesizing Thoughts from Activity 2: How should the public budget be allocated?

Algorithms vs Humans

- Arguments for algorithmic decision-making
  - *Potential to outperform humans in terms of accuracy and fairness*
    - They can leverage more data and potentially limitless computational power
  - *Potential to often be more transparent than humans*
    - Even if decisions are made using a black-box ML algorithm, being able to query the decisions in hypothetical scenarios makes it easier to assess fairness
  - *Potential to engage in deep mathematical reasoning about fairness*
    - Sometimes finding a fair outcome is an NP-hard problem
  - *Less bureaucracy, freeing up human time for other activities*
  - *…*

Algorithms vs Humans

- Arguments against algorithmic decision-making
  - *Algorithm may be designed to optimize the wrong objectives*
    - E.g. a social media platform designed to maximize the number of clicks rather than meaningful social connections, optimizing short-term objectives versus long-term goals
  - *Algorithms can often be less transparent than humans*
    - A black-box ML algorithm can be less transparent than a human following a well-documented and simple decision-making rule
  - *Being bound by a mathematical definition of fairness can be harmful*
    - No single definition may capture all facets of fairness in a context
  - *Potentially high energy consumption, impact on climate*
  - …

Algorithms vs Humans

- Poll 1
  - Suppose you are the mayor of Utopia City
  - Having heard of the amazing success of PB, you wish to conduct one
  - *If there are any complaints, you will be held accountable*
  - You have to choose between three systems for decision-making:
    1. A black-box machine learning algorithm, which can be trained to optimize any mathematically well-defined objectives
    2. A committee of city officials
    3. A committee of residents (citizen's assembly)
  - All three systems will try to optimize the same high-level goals and neither is fully transparent
  - Which system would you choose?

Algorithms vs Humans

- Poll 2
  - Consider the same problem, but now you're a resident of Utopia City
  - *You want to make sure that your voice is heard, the funds are allocated fairly and efficiently, and your neighborhood gets its deserved share of the funding*
  - You are given the option to provide your preference between the same three systems:
    1. A black-box machine learning algorithm, which can be trained to optimize any mathematically well-defined objectives
    2. A committee of city officials
    3. A committee of residents (citizen's assembly)
  - Again, all three systems will try to optimize the same high-level goals and neither is fully transparent
  - Which system would you prefer?

Concluding Remarks

- Improving algorithmic decision-making systems
  - Improving the quality and diversity of data sources
  - Causal inferences to determine which factors truly affect the decision at hand
  - Regulations and audits
  - Ensuring diverse ideas are represented within the designers of algorithmic decision-making systems

- Future challenges
  - Using algorithms to aid and improve human decision-making
    - E.g., matching reviewers to papers in conference reviewing
    - Also, other ways to mix human and algorithmic decision-making
  - Real-time ethical decision-making, e.g., in self-driving cars
  - …