

Group Fairness in Prediction-Based Decision Making: From Moral Assessment to Implementation

Joachim Baumann^(1,2), Christoph Heitz⁽¹⁾

(1) Zurich University of Applied Sciences

(2) University of Zurich

Switzerland

Zurich University
of Applied Sciences



**School of
Engineering**

IDP Institute of Data Analysis
and Process Design

- Prediction-based decision systems may be unfair
- What is Fairness? – formal definition
- Step-by-step procedure for building fair decision systems
 - Moral assessment
 - Define appropriate fairness metrics
 - Implement fairness

Prediction-based decision making

Context: Decision-making on individuals

Prediction models are introduced for taking better decisions under uncertainty

Simplest case:

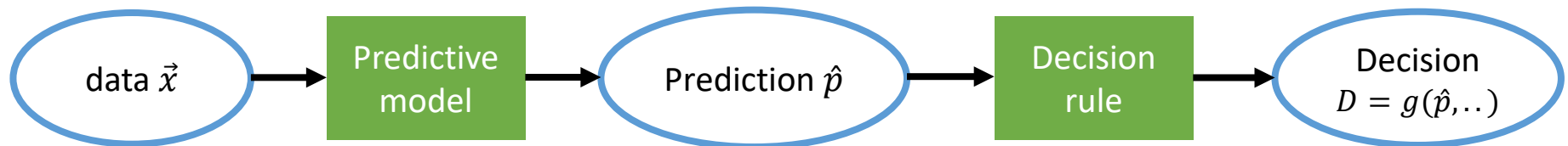
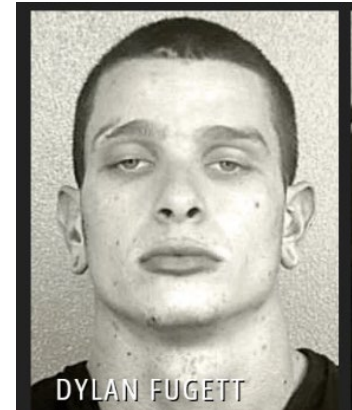
- Unknown variable $Y=0,1$
- Decision: $D=0,1$
- Theoretical best decision: If $Y=1$, then $D=1$. If $Y=0$, then $D=0$

Example: Releasing inmates on bail - yes or no?

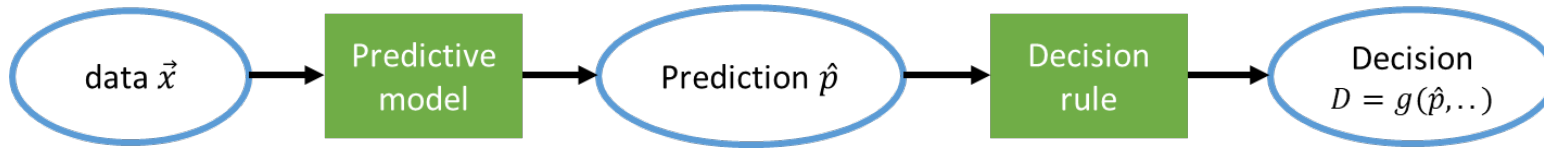
- $Y=1$: will reoffend, when released
- $D=1$: keep in prison

Implementation:

- Data \vec{x}
- Prediction model: $\hat{p} = P[Y = 1|\vec{x}]$
- Decision rule: threshold on \hat{p} - optimized for goal of decision maker



Fairness of prediction-based decision systems



Decision systems are build to take better decisions

But: At the same time, they create consequences on lives of individuals

- Different social groups \rightarrow different consequences?
- Unfairness = discrimination = systematic disadvantage of a social group

Example: COMPAS (Angwin et al. 2016)

- False Positive Rate different for Caucasians and African Americans

State of the art:

- Many incompatible fairness metrics suggested – what to do??
- Good decisions under fairness constraint? - only partial solutions available



Contribution of this paper

1. Step-by-step procedure how to build fair decision algorithms for a wide range of applications
 - Moral assessment: What does «fairness» mean in a given context?
 - Measuring fairness: Find the appropriate mathematical representation (fairness metrics)
 - Building fair algorithms: Solving the mathematical problem

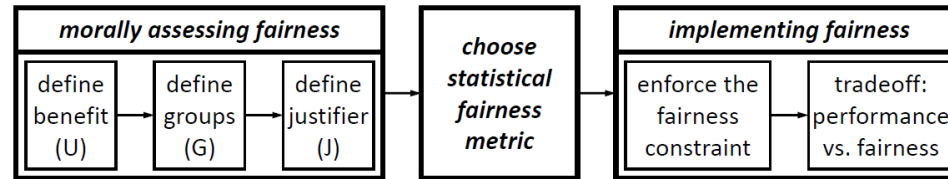
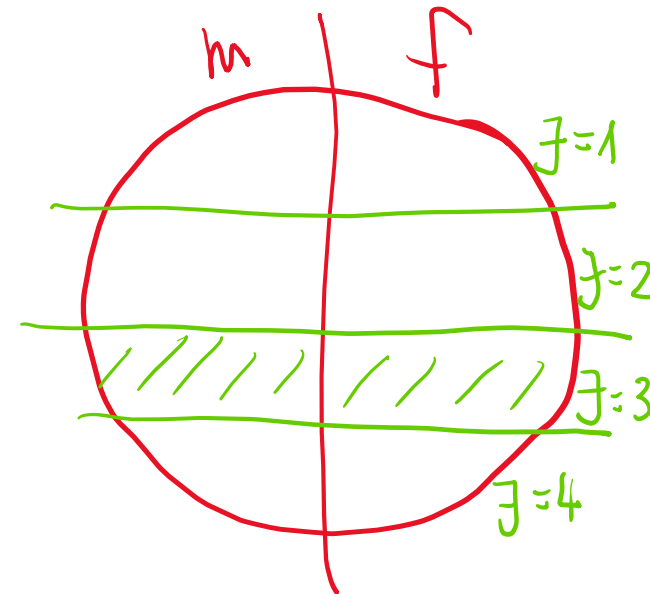
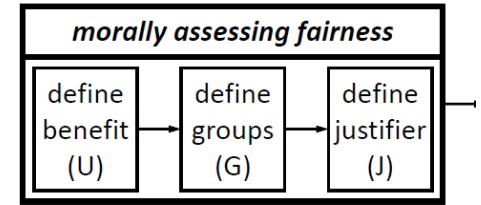


Fig. 1. A step-by-step procedure for fair prediction-based decision making

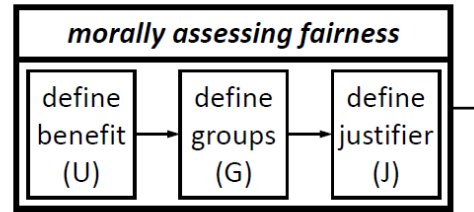
2. Extensions of current state-of-the art
 - Known fairness principle (Fairness of Equality) extended to all established fairness metrics
 - Dealing with morally induced relaxations of fairness metrics

What is Fairness: The principle of «Fair Equality of Chances»

- Any decision system leads to **distribution of benefit/harm** among individuals
 - Distributive justice as basic philosophical concept
 - Benefit is function of Y and D
- **Decision subjects (DS)**: representatives of socially salient groups
 - E.g. «men», «women»
- **Utility of Decision subjects (U_{DS})**: how much benefit for a decision subject of a group?
 - Focus on systematic effects: **expected** utility $E(U_{DS})$, averaged over group
- **Fair Equality of Chances (FEC)** (Loi et al, 2019):
Basic Idea: **$E(U_{DS})$ should be equal for groups**
- If **different moral claims** within population:
 - Population is split into subpopulations $J=1,2,\dots$ with the same moral claim
 - FEC means: $E(U_{DS}|J = j, m) = E(U_{DS}|J = j, f)$, for all relevant j
 - J is called “justifier”



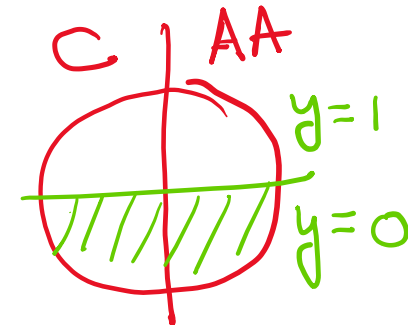
Step 1: moral assessment



- What is „benefit“?
- Which groups to compare?
- Define justifiers
 - Fairness only considered on subpopulations with the same value of J
 - Define which values j^* of J are morally relevant



- Benefit = being released on bail
- „Caucasian“ vs. „African American“
- Different moral claims to freedom
 - $Y=1$ (will re-offend): lower claim
 - $Y=0$ (will not re-offend): higher claim



Step 2: Find appropriate group fairness metric

Shown: established group fairness metrics are equivalent to specific choices

- what is considered the benefit (as a function of Y and D)
- what is considered a morally valid justifier J
- Which values j^* of J are considered morally relevant

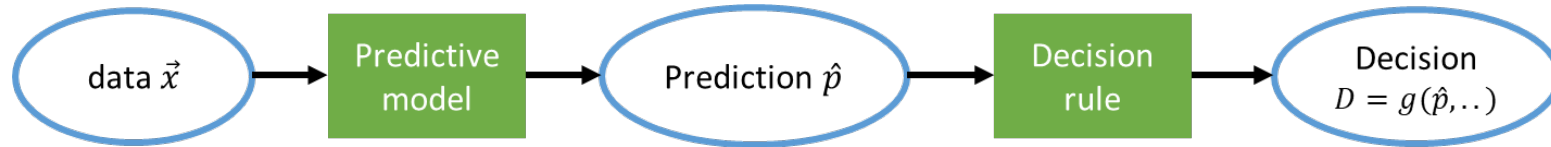
TABLE I

CHOOSING THE MOST APPROPRIATE GROUP FAIRNESS CRITERIA FOR A GIVEN CONTEXT ACCORDING TO THE E-FEC PRINCIPLE W.R.T. $G = \{m, f\}$

Group fairness metric	U_{DS}	J	j^*	Mathematical representation
Independence ^a	D	-	-	$P(D = 1 G = m) = P(D = 1 G = f)$
Conditional statistical parity	D	L	-	$P(D = 1 L = l, G = m) = P(D = 1 L = l, G = f)$
Separation ^b	-	-	-	$P(D = 1 Y = i, G = m) = P(D = 1 Y = i, G = f), i \in \{0, 1\}$
TPR parity ^c	D	Y	{1}	$P(D = 1 Y = 1, G = m) = P(D = 1 Y = 1, G = f)$
FPR parity ^d	-	-	{0}	$P(D = 1 Y = 0, G = m) = P(D = 1 Y = 0, G = f)$
Sufficiency ^e	-	-	-	$P(Y = 1 D = j, G = m) = P(Y = 1 D = j, G = f), j \in \{0, 1\}$
PPV parity ^f	Y	D	{1}	$P(Y = 1 D = 1, G = m) = P(Y = 1 D = 1, G = f)$
FOR parity	-	-	{0}	$P(Y = 1 D = 0, G = m) = P(Y = 1 D = 0, G = f)$

Equivalent metrics: ^astatistical parity, demographic parity, group fairness, ^bequalized odds, disparate mistreatment, ^cequality of opportunity, false negative error rate balance, ^dfalse positive error rate balance, predictive equality, ^econditional use accuracy equality, ^fpredictive parity, outcome test

Step 3: Implement fairness in decision making



Implementation of fairness by optimizing the decision rule:

- Mathematical formulation:

$$E(U_{DM}) = \max, \text{ subject to FEC constraint}$$

- Solution: Optimum decision rules are always group-dependent threshold rules, applied to \hat{p}
 - Hardt et al. (2016) for separation, Corbett-Davies et al (2017) for independence, Baumann and Heitz (2022) for sufficiency
- Hard constraints can be softened for allowing a transition between no fairness and full fairness (see paper)

Conclusion: how to build fair decision algorithms?

- Fairness: How is benefit distributed among socially salient groups?
- Principle of Fair Equality of Chances (FEC) extended
 - All classical fairness metrics of Fair ML literature are special cases
- Three-step procedure
 - Moral assessment
 - Choice of appropriate fairness metric
 - Solving a constraint optimization problem
- General approach - can be applied to all practical applications

Additional resources

- Python code for determining optimal threshold rules available at <https://github.com/joebaumann/fair-prediction-based-decision-making>
- Baumann, J., Hannák, A., and Heitz, C. (June 2022): Enforcing Group Fairness in Algorithmic Decision Making: Utility Maximization Under Sufficiency, Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT).
- Hertweck, C., Baumann, J., Loi, M., Viganò, E., and Heitz, C. (June 2022): A Justice-Based Framework for the Analysis of Algorithmic Fairness-Utility Trade-Offs, <http://arxiv.org/abs/2206.02891>
- Baumann, J., Hertweck, C., Loi, M., and Heitz, C. (June 2022): Distributive Justice as the Foundational Premise of Fair ML: Unification, Extension, and Interpretation of Group Fairness Metrics. <http://arxiv.org/abs/2206.02897>

References:

- See paper