Fairness in Learning: Classic and Contextual Bandits

M. Joseph, M. Kearns, J. Morgenstern, A. Roth
{majos, mkearns, jamiemor, aaroth}@cis.upenn.edu

High-Level Motivation
- Machine learning can be unfair in many ways: data that encodes existing biases; data collection feedback loops; different populations having different properties; less data about minority populations . . .
- How do we define “fair learning”?
- What is the performance cost of being fair?

General Problem Setting
- We study the bandits setting: \( k \) arms, on day \( t \in \mathbb{T} \) choose arm \( i_t \) and observe noisy reward \( r_{ti}^{t} \)
- Goal: maximize \( \sum_{t=1}^{T} r_{ti}^{t} \), measure performance by regret \( R(T) = \sum_{t=1}^{T} \left( \max_{i \in [k]} E[r_{ti}^{t}] - r_{ti}^{t} \right) \)
- Models a program that learns to grant loans to \( k \) different groups by granting loans to one member of one group each day

General Fairness Definition
- Algorithm \( \mathcal{A} \) is fair if with probability \( \geq 1 - \delta \), for all days \( t \in \mathbb{T} \) and for all \( i, j \in [k] \)
  \[ E[r_{ti}^{t}] \geq E[r_{tj}^{t}] \Rightarrow \pi_{1|hi_{1},...,h_{t-1}}^{t} \geq \pi_{1|jh_{1},...,h_{t-1}}^{t} \]
  where \( \pi_{1|hi_{1},...,h_{t-1}}^{t} = \mathbb{P}[\text{choose } i \text{ in round } t \text{ after observing } h_{1}, \ldots, h_{t-1}] \)
- “With high probability, never more likely to choose a worse arm than a better arm”

Why is Fairness Hard?
- Optimal policies always play the expected best arm and therefore are fair. Challenge: how to learn the optimal policy fairly?

Classic Bandits Setting
- \( \mu_i \) for each arm \( i \) such that for all \( i \) and \( t \) \( E[r_{ti}^{t}] = \mu_i \)
- Fair: \( \mu_i \geq \mu_j \Rightarrow \pi_{1|hi_{1},...,h_{t-1}}^{t} \geq \pi_{1|jh_{1},...,h_{t-1}}^{t} \)
- “With high probability, never more likely to choose an arm with lower \( \mu \) than an arm with higher \( \mu \)”

A Fair Classic Bandit Algorithm: FairBandits
- Uses confidence intervals around estimated means to reason about relative quality; fairness forces chaining
- In round \( t \): pick uniformly at random from “chain” of top arms (top connected component of overlapping confidence intervals)

Cost of Fairness in Classic Bandits
- FairBandits regret upper bound \( R(T) = \tilde{O}(\sqrt{k^3 T}) \)
- Regret lower bound (any fair algorithm) \( R(T) = \Omega(k^3) \), while \( R(T) = \tilde{O}(\sqrt{k T}) \) (unfair)

Contextual Bandits Setting
- Function \( f_i \in \mathbb{C} \) for \( i \in [k] \), \( x_t^i \in \mathbb{R}^d \) for \( t \in \mathbb{T} \), \( i \in [k] \) such that \( E[r_{ti}^{t}] = f_i(x_t^i) \)
- Fair: \( f_i(x_t^i) \geq C f_j(x_t^i) \Rightarrow \pi_{1|hi_{1},...,h_{t-1}}^{t} \geq C \pi_{1|jh_{1},...,h_{t-1}}^{t} \)
- “With high probability, never more likely to choose an arm with lower \( f(x^i) \) than an arm with higher \( f(x^i) \)”

Fair Contextual Bandits and KWIK Learning
- \( C \) is KWIK-learnable [1] with poly KWIK bound \( \leftrightarrow C \) can be learned fairly with poly regret
- For \( d \)-dimensional linear functions, KWIK bounds [2] imply fair learning with \( R(T) = \tilde{O}(\max \{ T^{4/5} k^{6/5} d^{3/5}, k^3 \}) \)
- For \( d \)-dimensional conjunctions, KWIK bounds [3] imply that no fair learning algorithm has a worst-case regret bound better than \( R(T) = \Omega(2^d) \)

References