Beyond Distributive Fairness in Algorithmic Decision Making:

Feature Selection for Procedurally Fair Learning

Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P. Gummadi and Adrian Weller
Algorithmic Decision Making

- Algorithms help people make decisions about
  - Hiring
  - Assigning social benefits
  - Granting bail

Are these algorithms fair?
Types of Algorithmic Fairness

<table>
<thead>
<tr>
<th>Distributive Fairness</th>
<th>Procedural Fairness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairness of <strong>decision making outcomes</strong></td>
<td>Fairness of the <strong>decision making process</strong></td>
</tr>
</tbody>
</table>

**Example**

- **Equal misclassification rates**
  - *Grant bail to high risk* white defendants
  - *Deny bail to low risk* black defendants
  → unfair outcomes

**Example**

- **Fairness of using features**
  - *Is it fair to use a feature in decision making?*
This Talk

• A **Notion** of Procedural Fairness: **Feature Usage Fairness**

• **Quantifying** Feature Usage Fairness

• **Mechanisms** for Achieving Feature Usage Fairness
Is it fair to use a feature?

- Discrimination
  - Sensitive *race, gender* vs non-sensitive features

- Fairness beyond discrimination
  - Volitionality (e.g., criminal history of defendant’s father)
    - Does the feature represent the result of volitional (i.e., voluntarily chosen) decisions made by the individual (e.g., number of prior offenses); or rather is it the result of circumstances beyond their control?
Is it fair to use a feature?

- Discrimination
  - Sensitive (**race**, **gender**) vs non-sensitive features

- Fairness beyond discrimination
  - Volitionality (e.g., criminal history of defendant’s father)
  - Relevance (e.g., defendant’s education)
    - Is the feature causally related or not to the decision outcomes?
Is it fair to use a feature?

• Discrimination
  • Sensitive (race, gender) vs non-sensitive features

• Fairness beyond discrimination
  • Volitionality (e.g., criminal history of defendant’s father)
  • Relevance (e.g., defendant’s education)
  • Reliability
    • How reliably can a feature be assessed (e.g., in credit assessments, opinions towards bankruptcy may be harder to reliably assess than number of prior bankruptcies)
Is it fair to use a feature?

- Discrimination
  - Sensitive (race, gender) vs non-sensitive features

- Fairness beyond discrimination
  - Volitionality (e.g., criminal history of defendant’s father)
  - Relevance (e.g., defendant’s education)
  - Reliability
  - Privacy
    - Does use of the feature give rise to a violation of the individual’s privacy?
Is it fair to use a feature?

- Discrimination
  - Sensitive (*race, gender*) vs non-sensitive features
- Fairness beyond discrimination
  - Volitionality (e.g., criminal history of defendant’s father)
  - Relevance (e.g., defendant’s education)
  - Reliability
  - Privacy

Background knowledge on fairness of features not in the data!
- Gather human moral judgments
Human Judgments of Fairness

- **Case study**: COMPAS tool for predicting criminal risk

Diagram:
- # Prior Offenses
- Charge Description
- Charge Degree
- # Juvenile Felonies
- # Juvenile Misd.
- # Juvenile Other
- Age
- Sex
- Race

- Directly related
- Distantly related
- Volitional
- Unrelated / Non-volitional / Physiological
Reasoning About Fairness

- What determines people’s moral judgments about fairness?

- There is **more to fairness than discrimination!**
Quantifying Fairness of a Classifier

- Feature usage fairness
  - The fraction of people that consider using that feature fair

- Feature usage fairness of a classifier
  - Fraction of people that consider all of its features fair
Fairness – Accuracy Tradeoff

- Intuitively
  - Adding features: higher accuracy, lower fairness
  - Removing features: lower accuracy, higher fairness

- There is a tradeoff between feature usage fairness & accuracy
Fair Feature Selection

• We want to select a subset of features that leads to
  • High accuracy
  • High feature usage fairness

• Formulation

\[
\begin{align*}
\text{maximize} & \quad \text{accuracy}(S) \\
\text{subject to} & \quad \text{unfairness}(S) \leq t
\end{align*}
\]

• How do we do this?
Naïve Approach

- Brute force
  - Train $2^n$ classifiers, $n =$ number of features

- Optimal Solution
- Not scalable! 30 features = more than 1 billion classifiers
- Is there an efficient alternative?
Submodular Optimization

- Feature usage unfairness is **submodular & monotone**
Fairness Properties - Monotonicity

• Feature unfairness is \textbf{monotone non-decreasing}

• \textit{Intuition}
  • A set function is monotone non-decreasing if \textbf{adding elements to a set cannot decrease its value}

• \textit{Definition}

\[
g(F_i \cup \{f\}) \geq g(F_i), \\
\forall F_i \subseteq F, f \in F \setminus F_i
\]
• Feature unfairness is submodular

• Intuition
  • A set function is submodular if it exhibits **diminishing marginal returns**

• Definition

\[
g(F_A \cup \{f\}) - g(F_A) \geq g(F_B \cup \{f\}) - g(F_B),
\]

\[
F_A \subseteq F_B \subseteq F, f \in F \setminus F_B
\]
Submodular Optimization

- Feature usage unfairness is **submodular & monotone**
- Submodular cost submodular knapsack problem
  - Approximate using **ISK** algorithm (Iyer and Bilmes, NIPS 2013)

- Efficient & scalable approximation
- Near optimal results
ISK algorithm

\[
\begin{align*}
\text{maximize} & \quad \text{accuracy}(S) \\
S \subseteq \mathcal{F} \quad \text{subject to} & \quad \text{unfairness}(S) \leq t
\end{align*}
\]

- Maps to **Submodular Cost Submodular Knapsack** problem

- **Guarantees & hardness**

  \[
  \text{Approx. factor}^* \\
  \text{Bi-Criterion factor}^# \\
  \left[1 - e^{-1}, \frac{K_f}{1 + (K_f - 1)(1 - \kappa_f)}\right]^#
  \]

- **Algorithm**
  - Iteratively finding modular approximations of submodular functions & solving the resulting knapsack problems
Accuracy Properties

- Accuracy is \textit{weakly submodular}

- \textit{More precisely}
  - Logistic loss with l2 regularization exhibits restricted strong convexity, which implies it’s weakly submodular

- \textit{Intuition on why this approach performs well}
  - \textit{Greedy algorithms preform well in practice} for logistic loss with l2 regularization
Procedural vs Distributive Fairness

• In the ProPublica COMPAS dataset:

  high process fairness $\rightarrow$ high outcome fairness
Key Points

• A Notion of Procedural Fairness: Feature Usage Fairness
  • Relies on people’s moral judgments
  • Beyond discrimination: volitionality, relevance, reliability...

• Quantifying Feature Usage Fairness of a Decision Making System
  • Fraction of people that consider all features fair

• Mechanisms for Achieving Feature Usage Fairness
  • Control tradeoffs between fairness and accuracy
  • Submodular measure → scalable fair feature selection