

Fairness-enhancing interventions in stream classification

V. Iosifidis¹ H.T. Thi Ngoc¹ E. Ntoutsi¹

¹University of Leibniz
Hannover, Germany

DEXA Conference, October 4, 2019

Outline

- 1 Bias Everywhere
- 2 Can algorithms be biased ?
 - Stream Classification: Concept Drifts and Fairness ?
- 3 Proposed Framework
 - Architecture
 - Discrimination Measure
 - Chunk Massaging
 - Chunk Re-weighting
 - Concept drift strategies
- 4 Evaluation
 - Baselines
 - Datasets
 - Results
- 5 Summary

Nanny or Wife ? ¹



¹<http://www.bbc.com/news/world-asia-39244325>

The Meeting Denial ²

Student Race and Gender	Emails Ignored		Meetings Denied	
	%	% Increase Relative to Caucasian Males	%	% Increase Relative to Caucasian Males
Caucasian Male	26.5%	N/A	52.4%	N/A
Caucasian Female	29.8%	12.5%	52.9%	1.1%
Black Male	32.5%	22.6%	61.3%	17.0%
Black Female	34.4%	29.8%	60.0%	14.6%
Hispanic Male	36.9%	39.2%	58.2%	11.1%
Hispanic Female	27.1%	2.3%	55.7%	6.3%
Indian Male	41.8%	57.7%	68.2%	30.2%
Indian Female	37.7%	42.3%	67.9%	29.7%
Chinese Male	36.7%	38.3%	66.8%	27.6%
Chinese Female	46.9%	77.0%	62.9%	20.2%

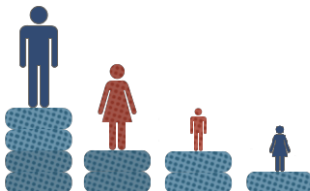
Copyright Katherine Milkman, Modupe Akinola and Dolly Chugh 2012

²<http://knowledge.wharton.upenn.edu/article/e-mails-ignored-meetings-denied-bias-at-the-search-stage-limits-diversity>

Can algorithms be biased?

- Decision support systems are data-driven.
- Decision rules are generated by data patterns.
- Data is often produced by humans.
- Algorithms can reinforce human prejudices.

“If data contains bias then algorithms which are trained on this data will also produce biased results.”



Low or High Risk ? ³

Two Petty Theft Arrests

<p>VERNON PRATER</p> <hr/> <p>Prior Offenses 2 armed robberies, 1 attempted armed robbery</p> <hr/> <p>Subsequent Offenses 1 grand theft</p>	<p>BRISHA BORDEN</p> <hr/> <p>Prior Offenses 4 juvenile misdemeanors</p> <hr/> <p>Subsequent Offenses None</p>
<p>LOW RISK 3</p>	<p>HIGH RISK 8</p>

Two Drug Possession Arrests

<p>DYLAN FUGETT</p> <hr/> <p>Prior Offense 1 attempted burglary</p> <hr/> <p>Subsequent Offenses 3 drug possessions</p>	<p>BERNARD PARKER</p> <hr/> <p>Prior Offense 1 resisting arrest without violence</p> <hr/> <p>Subsequent Offenses None</p>
<p>LOW RISK 3</p>	<p>HIGH RISK 10</p>

³<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

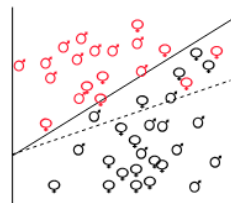
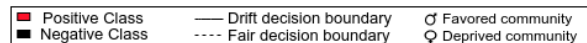
Fairness in Streams

Stream Classification

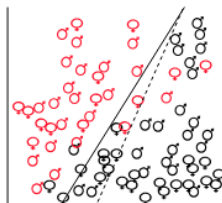
Classification boundary changes as a result of model adaptation

Fairness-aware Stream Classification

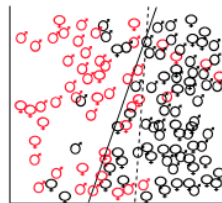
Classification boundary **MUST** consider fair outcomes!



t = 1



t = 2



t = 3

Framework's Overview



Discrimination Measure

Basic Notation for Fairness

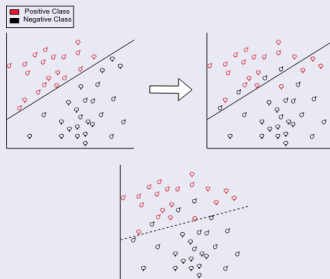
Sensitive Attribute SA	(predicted) class	
	Rejected	Granted
s (Female)	DR_t (<i>deprived rejected</i>)	DG_t (<i>deprived granted</i>)
\bar{s} (Male)	FR_t (<i>favored rejected</i>)	FG_t (<i>favored granted</i>)

Statistical Parity

$$disc_S(F, S_t) = \frac{FG_t}{FG_t + FR_t} - \frac{DG_t}{DG_t + DR_t} \quad (1)$$

Chunk Massaging

Method

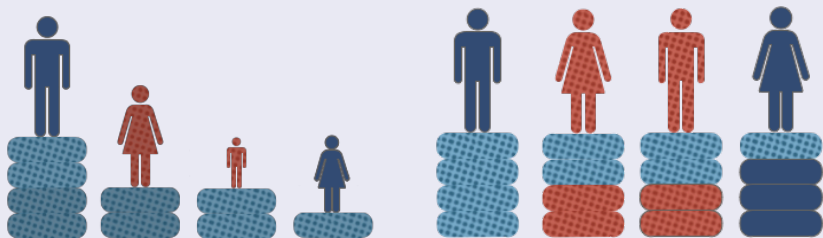


Number of swaps

$$M_t = \frac{FG_{S_t} * (DG_{S_t} + DR_{S_t}) - DG_{S_t} * (FG_{S_t} + FR_{S_t})}{|S_t|} \quad (2)$$

Chunk Re-weighting

Method



Weight Estimation

$$W_t^{FG} = \frac{|\bar{S}_t| * |\{x \in S_t | (x.C = \text{"granted"})\}|}{|S_t| * |FG_{S_t}|} \quad (3)$$

Framework Strategies

Model Adaptation Strategies

- **Accum&FullTrain**: Keep history and update per original **or** corrected chunk.
- **Reset&FullTrain**: Discard history and update per original **or** corrected chunk.
- **Accum&CorrectedTrain**: Keep history and update **per corrected chunk**.
- **Reset&CorrectedTrain**: Discard history and update **per corrected chunk**.

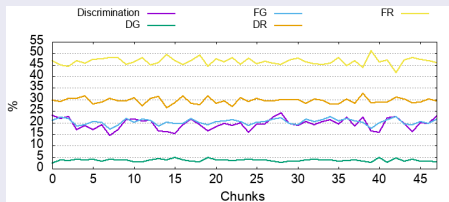
Baselines

Methods

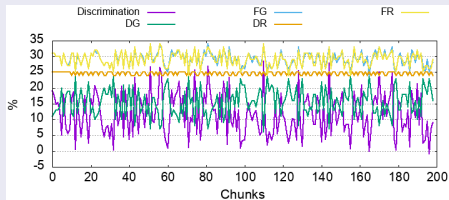
- **B.NoSA** (Baseline NoSensitiveAttribute): The classifier F does not employ SA neither in training nor in testing.
- **B.RESET** (Baseline Reset): Discard history in case of discrimination.

Datasets

Adult Census dataset

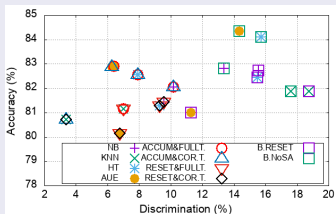


Synthetic dataset

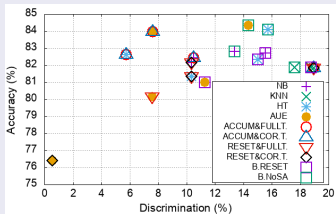


Results: Adult Census dataset

Massaging

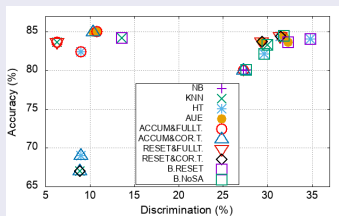


Re-weighting

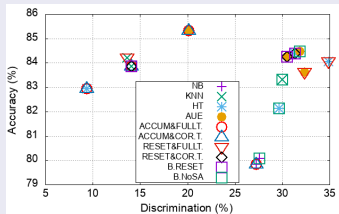


Results: Synthetic dataset

Massaging



Re-weighting



Summary

- Our framework is model agnostic.
- Our framework tackles discrimination and maintains relatively good performance.
- Massaging performs better than re-weighting.
- **Future work:** Deal with unfair outcomes in imbalanced streams.

Thanks.

Questions?

Contact: {iosifidis,ntoutsis}@L3S.de