IsyE 8813: Algorithmic Foundations of Ethical ML

Fairness Reading

Add yourself to the schedule by November 18th
Presentations on November 30th and December 2nd
Summaries due December 9th

Summarize 1 paper on fairness if working alone, x if working in a group of x people. For each paper, please i) provide context and motivation for the studied problem, ii) summarize the main results of the paper, iii) provide an overview of the key technical tools use to achieve these results, and iv) find a few questions left open by the paper.

If you are summarizing 2 papers, please try to pick papers that are related to each other, and to discuss how these papers interact with each other (is one a follow-up to the other? Do they complete each other? Do they study the same problem under different assumptions? Do they use the same techniques but apply them to different problems?). Feel free to pick topics or papers of interest that are not included in the list below.

Your paper summary should be emailed to jziani3@gatech.edu. You should also prepare a 15 to 20-minute class presentation about the paper you have read if working alone, and a 30 to 40 minute presentation if working in a group of 2.

List of Topics

Remember that the lists below are non-exhaustive. Feel free to pick papers outside of the ones proposed below.

Any fairness-related paper from recent workshops and conferences A few possible conferences:

- Symposium on the Foundations of Responsible Computing (FORC)
- Conference on Fairness, Accountability, and Transparency (FAccT)
- Workshop on Mechanism Design for Social Good (MD4SG)
- Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)
- Conference on Artificial Intelligence, Ethics, and Society (AIES)

Basic definitions of fairness and interventions for fairness Overlaps with the topics we saw in class, but provides more details.

- Fairness Through Awareness. Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, Rich Zemel.
- Equality of Opportunity in Supervised Learning. Moritz Hardt, Eric Price, Nathan Srebro.
- Inherent Trade-Offs in the Fair Determination of Risk Scores. Jon Kleinberg, Sendhil Mullainathan, Manish Raghavan.
- Certifying and removing disparate impact. Michael Feldman, Sorelle Friedler, John Moeller, Carlos Scheidegger, Suresh Venkatasubramanian.
- Calibration for the (Computationally-Identifiable) Masses. Ursula Hébert-Johnson, Michael P. Kim, Omer Reingold, Guy N. Rothblum.
- Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness. Michael Kearns, Seth Neel, Aaron Roth, Zhiwei Steven Wu.
- Decoupled classifiers for fair and efficient machine learning. Cynthia Dwork, Nicole Immorlica, Adam Tauman Kalai, Max Leiserson.
- Equalized odds postprocessing under imperfect group information. Pranjal Awasthi, Matthäus Kleindessner, Jamie Morgenstern.
- Algorithmic decision making and the cost of fairness. Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, Aziz Huq.

Long-term fairness What happens if the decision we make today may have an impact in the long-term? What if feedback loops are present, and the decisions we make today affect the populations we face tomorrow?

- Delayed Impact of Fair Machine Learning. Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt
- A Short-term Intervention for Long-term Fairness in the Labor Market. Lily Hu, Yiling Chen.
- On the long-term impact of algorithmic decision policies: Effort unfairness and feature segregation through social learning.
- Fair Algorithms for Learning in Allocation Problems. Hadi Elzayn, Shahin Jabbari, Christopher Jung, Michael Kearns, Seth Neel, Aaron Roth, Zachary Schutzman.

- Runaway Feedback Loops in Predictive Policing. Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, Suresh Venkatasubramanian.
- Allocating Opportunities in a Dynamic Model of Intergenerational Mobility. Hoda Heidari, Jon Kleinberg.
- From Fair Decision Making to Social Equality. Hussein Mozannar, Mesrob I. Ohannessian, Nathan Srebro.

Fairness in composed decisions What happens if we do not look at single algorithm and try to make it fair, but rather we have to deal with a pipeline of several decisions?

- Fairness Under Composition. Cynthia Dwork, Christina Ilvento.
- Individual Fairness in Pipelines. Cynthia Dwork, Christina Ilvento, Meena Jagadeesan.
- Downstream Effect of Affirmative Action. Sampath Kannan, Aaron Roth, Juba Ziani.
- Pipeline Interventions. Eshwar Ram Arunachaleswaran, Sampath Kannan, Aaron Roth, Juba Ziani.
- Fair Pipelines. Amanda Bower, Sarah N Kitchen, Laura Niss, Martin J Strauss, Alexander Vargas, and Suresh Venkatasubramanian.

Fairness under Strategic Behavior What if the agents we make decisions on are strategic, and try to adapt to the classifier or decision rule?

- The Social Cost of Strategic Classification. Smitha Milli, John Miller, Anca D. Dragan, Moritz Hardt.
- The Disparate Effects of Strategic Manipulation. L Hu, N Immorlica, JW Vaughan.
- Actionable Recourse in Linear Classification. Berk Ustun, Alexander Spangher, Yang Liu.
- Fairness Interventions as (Dis)Incentives for Strategic Manipulation. Xueru Zhang, Mohammad Mahdi Khalili, Kun Jin, Parinaz Naghizadeh, Mingyan Liu.

Fairness and Privacy Are fairness and privacy compatible? How do they work with each other?

- On the Compatibility of Privacy and Fairness. Rachel Cummings, Varun Gupta, Dhamma Kimpara, Jamie Morgenstern.
- Differential Privacy Has Disparate Impact on Model Accuracy. Eugene Bagdasaryan, Omid Poursaeed, Vitaly Shmatikov.

- Decision Making with Differential Privacy under a Fairness Lens. Ferdinando Fioretto, Cuong Tran, Pascal Van Hentenryck.
- Differentially Private Fair Learning. Matthew Jagielski, Michael Kearns, Jieming Mao, Alina Oprea, Aaron Roth, Saeed Sharifi-Malvajerdi, Jonathan Ullman.

Individual Fairness and Metrics How to deal with unknown metrics for individual fairness? How to learn the metric from auditors, or guarantee that we do not violate the metric too much/too often?

- Online learning with an unknown fairness metric. Stephen Gillen, Christopher Jung, Michael J. Kearns, and Aaron Roth.
- Metric-Free Individual Fairness in Online Learning. Yahav Bechavod, Christopher Jung, Zhiwei Steven Wu.
- Metric Learning for Individual Fairness. Christina Ilvento.
- An Algorithmic Framework for Fairness Elicitation. Chistopher Jung, Michael Kearns, Seth Neel, Aaron Roth, Logan Stapleton.

Fairness in Online Decision-Making

- Metric-Free Individual Fairness in Online Learning. Yahav Bechavod, Christopher Jung, Zhiwei Steven Wu.
- Fairness in learning: Classic and contextual bandits. Matthew Joseph, Michael Kearns, Jamie Morgenstern, Aaron Roth.
- Meritocratic fairness for infinite and contextual bandits. Matthew Joseph, Michael J. Kearns, Jamie Morgenstern, Seth Neel, and Aaron Roth.
- Online learning with an unknown fairness metric. Stephen Gillen, Christopher Jung, Michael J. Kearns, and Aaron Roth.
- Individual fairness in hindsight. Swati Gupta and Vijay Kamble.
- Fairness Incentives for Myopic Agents. Sampath Kannan, Michael Kearns, Jamie Morgenstern, Mallesh Pai, Aaron Roth, Rakesh Vohra, Z. Steven Wu.

Some application areas: Ad auctions:

- Fairness in ad auctions through inverse proportionality. Shuchi Chawla, Meena Jagadeesan.
- Multi-Category Fairness in Sponsored Search Auctions. Shuchi Chawla, Christina Ilvento, and Meena Jagadeesan.

- Toward Controlling Discrimination in Online Ad Auctions. L. Elisa Celis, Anay Mehrotra, and Nisheeth K. Vishnoi.

College Admissions and Hiring:

- Dropping Standardized Testing for Admissions: Differential Variance and Access. Nikhil Garg, Hannah Li, and Faidra Monachou.
- Downstream Effects of Affirmative Action. Sampath Kannan, Aaron Roth, Juba Ziani.
- A Short-term Intervention for Long-term Fairness in the Labor Market. Lily Hu, Yiling Chen.

Predictive Policing:

- Runaway Feedback Loops in Predictive Policing. Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, Suresh Venkatasubramanian.
- To predict and serve? Kristian Lum and William Isaac.

Etc.