VECTOR INSTITUTE | INSTITUT VECTEUR

Endless Summer School

Privacy and Fairness

Wednesday 18th April 2018

Agenda

8:30am – Registration and Coffee

9:00am – Introduction and Opening Remarks

9:15am – Richard Zemel, “Ensuring Fair and Responsible Automated Decisions”

10:30am – Break

10:45am – Sasho Nikolov, “Algorithmic Techniques in Differential Privacy”

12:00pm – Lunch and Networking

1:00pm – Andrei Fajardo, “Differential Privacy in Practice”

1:30pm – Pamela Snively, “Respecting Privacy – From Consent to Accountability”

2:00pm – Carole Piovesan, “AI and the Law: Exploring the Legal Implications of Artificial Intelligence”

2:30pm – Break

2:45pm – Elliot Creager, “Learning Adversarially Fair and Transferable Representations”

3:10 – David Madras, “Predict Responsibly: Increasing Fairness by Learning to Defer”

3:35 – Angus Galloway, “How Does Generalization Relate to Privacy?”

4:00 – Geoff Roeder, “Generative Models for Automatic Design”

4:30pm – Closing Remarks

Presentations

in order of appearance

Ensuring Fair and Responsible Automated Decisions

Information systems are becoming increasingly reliant on statistical inference and learning to render all sorts of decisions, including the issuing of bank loans, the targeting of advertising, and the provision of health care. This growing use of automated decision-making has sparked heated debate among philosophers, policy-makers, and lawyers, with critics voicing concerns with bias and discrimination. Bias against some specific groups may be ameliorated by attempting to make the automated decision-maker blind to some attributes, but this is difficult, as many attributes may be correlated with the particular one. The basic aim then is to make fair decisions, i.e., ones that are not unduly biased for or against specific subgroups in the population. I will discuss various computational formulations and approaches to this problem.

Richard Zemel is a Professor of Computer Science at the University of Toronto, where he has been a faculty member since 2000, and Research Director of the Vector Institute. Prior to that, he was an Assistant Professor in Computer Science and Psychology at the University of Arizona and a Postdoctoral Fellow at the Salk Institute and at Carnegie Mellon University. He received a BSc degree in History & Science from Harvard University in 1984 and a PhD in Computer Science from the University of Toronto in 1993. He is also the co-founder of SmartFinance, a financial technology start-up specializing in data enrichment and natural language processing.

Algorithmic Techniques in Differential Privacy

The practical applicability of data analysis to the sciences and to decision-making is often limited by privacy concerns. Without strong privacy guarantees, individuals may refuse to participate in data collection, or not provide truthful information. These issues limit the validity of the data analysis results, and make balancing privacy and usefulness essential in many applications of statistics and machine learning. Over the last ten years, differential privacy has emerged as a framework that enables data analysis while giving strong privacy guarantees. In this talk I will describe basic concepts and techniques in differential privacy, and survey some more recent advances.

Aleksandar (Sasho) Nikolov is an assistant professor at the University of Toronto. Sasho received his PhD from Rutgers University, where his supervisor was S. Muthukrishnan, and did a postdoc with the Theory Group at Microsoft Research in Redmond. He is a Canada Research Chair in Algorithms and Privacy, and in 2012-14 was a Simons Graduate Fellow in Computer Science.
Sasho is broadly interested in theoretical computer science and algorithms, and specifically in the foundations of data privacy, discrepancy theory, convex geometry, and their applications to computer science. He is also interested in sublinear and parallel algorithms for large data.

Differential Privacy in Practice

Differential privacy, a technique to insert noise into a machine learning or statistical model that obfuscates the relationship between individual data points and the insights they generate, is gradually shifting from theoretical research to practical applications. This talk will share lessons learned applying differential privacy to a model that segments customers based on propensity to purchase auto insurance.

Andrei Fajardo is a Machine Learning Scientist at Integrate AI. Before joining integrate, Andrei obtained a PhD in Statistics from the University of Waterloo.

Respecting Privacy – From Consent to Accountability

I will review the challenges with using the old privacy paradigms when dealing with the complexities of big data and AI. I will outline the current tensions and the work that is being done by leaders in the industry to address these challenges in a manner that will both respect privacy and support innovation.

Pamela Snively is TELUS' Chief Data & Trust Officer. She leads the team responsible for privacy and, more broadly, data governance. Pamela is charged with enabling and supporting TELUS' commitment to earning and maintaining customer trust through demonstrably accountable and ethical data practices.

Pamela also leads TELUS Wise, a free educational program that empowers Canadians to stay safe in our digital world. The program offers interactive and informative workshops and content to help Canadians have a positive experience as digital citizens. Topics include protecting online security, privacy, and reputation, rising above cyberbullying, and using technology responsibly. Pamela also supports

TELUS' commitment to integrity through her leadership of a variety of initiatives, including the Anti-bribery & Corruption and Competition Law programs.

Prior to joining TELUS, Pamela managed a privacy and data management consulting practice. As a lawyer and consultant, she provided a broad range of privacy, risk-management and compliance advice in the private, public and health sectors. Pamela formerly held Chief Privacy and Risk Officer roles in large business process outsourcing companies in Canada.

AI and the Law: Exploring the Legal Implications of Artificial Intelligence

This presentation will explore some of the key legal issues associated with artificial intelligence including explainability and accountability.

Carole Piovesan is a litigator at McCarthy Tetrault. She is the firm lead on artificial intelligence and an active member of the firm's Cybersecurity, Privacy and Data Management group. She co-authored the firm's White Paper entitled "*From Chatbots to Self-Driving Cars: The Legal Risks of Adopting Artificial Intelligence in Your Business*", and regularly writes and speaks on the topic of AI and the law.

Learning Adversarially Fair and Transferable Representations

We advocate for representation learning as the key to mitigating unfair prediction outcomes downstream. We envision a scenario where learned representations may be handed off to other entities with unknown objectives. We propose and explore adversarial representation learning as a natural method of ensuring those entities will act fairly, and connect group fairness (demographic parity, equalized odds, and equal opportunity) to different adversarial objectives. Through worst-case theoretical guarantees and experimental validation, we show that the choice of this objective is crucial to fair prediction. Furthermore, we present the first in-depth experimental demonstration of fair transfer learning, by showing that our learned representations admit fair predictions on new tasks while maintaining utility, an essential goal of fair representation learning.

Elliot Creager is a Ph.D. student at the University of Toronto and the Vector Institute for Artificial Intelligence. He motivates his research by first asking what else beyond predictive accuracy we should expect of our artificial intelligence systems (for example, equitable and interpretable predictions), then engineering solutions to these auxiliary goals. His supervisor is Richard Zemel.

Predict Responsibly: Increasing Fairness by Learning to Defer

In high-stakes ML applications, there are multiple decision-makers involved, both automated and human. The interaction between these agents often goes unaddressed in algorithmic development. In this work, we explore a simple version of this interaction with a two-stage framework containing an automated model and an external decision-maker. The model can choose to say IDK, and pass the decision downstream, as explored in rejection learning. We extend this concept by proposing learning to defer, which generalizes the rejection learning framework by considering the effect of the other agents in the decision-making process. We propose a learning algorithm which accounts for potential biases held by external decision-makers in a system. Experiments on real-world datasets demonstrate that learning to defer can make a system not only more accurate but also less biased. Even when operated by highly biased users, we show that deferring models can still greatly improve the fairness of the entire system.

David Madras is a PhD student in the Machine Learning Group at the University of Toronto and Vector Institute, supervised by Richard Zemel. He is interested in exploring new machine learning models and algorithms, particularly pertaining to fairness, privacy, and safety in machine learning, as well the role of AI in decision-making systems and the law.

How does generalization relate to privacy?

Learning models that generalize to novel data is the ultimate goal in machine learning. The adversarial examples phenomenon challenges the notion that the current instantiation of deep neural networks generalize well, or that they are aware of what they don’t know. I’ll introduce various threat models in the adversarial setting, demonstrate practical attacks, and explore limitations of using threat models to characterize robustness. I believe privacy, “interpretability”, and generalization can be satisfied concurrently, and will summarize our recent work bridging these areas.

Angus Galloway is an engineering graduate student at the University of Guelph with a research emphasis on robust machine learning, and defenses against adversarial examples. His research has implications for deep learning in performance-critical scenarios such as autonomous driving, and making medical decisions--where reliable confidence estimates are essential. His work on adversarial examples is to appear at ICLR 2018, with two related works submitted to ICML and UAI. He is also a contributor to the CleverHans library for benchmarking the robustness of machine learning models. Angus advises two startups at the intersection of hardware and machine learning, and has experience on several engineering teams in the semiconductor industry.

Generative Models for Automatic Design

Generative models can be used to produce designs that obey hard-to-specify constraints while still producing plausible examples. Recent examples of this include drug design, text with desired sentiment, or images with desired captions. However, most previous applications of generative models to design are based on bespoke, ad-hoc procedures. We give a unifying treatment of generative design based on probabilistic generative models. Some of these models can be trained end-to-end, can take advantage of both labelled and unlabelled examples, and automatically trade off between different design goals.

Geoffrey Roeder is a student researcher at the Vector Institute for Artificial Intelligence. His research areas include probabilistic generative models, a class of machine learning model that automatically discovers underlying patterns in data and uses them to generate new structured content.