Automated Algorithms and Risk: Two Sides of the Coin

Daniel B. Neill, Ph.D.
Associate Professor of Computer Science, Public Service, and Urban Analytics
New York University

E-mail: daniel.neill@nyu.edu
Web: http://www.cs.nyu.edu/~neill
The use of **automated algorithms** for decision making has become increasingly ubiquitous across a wide variety of fields...

- Online marketing
- Health care diagnosis & treatment
- Lending
- Hiring
- Policing & criminal justice
- Allocation of city services

Across all of these domains, the use of such algorithms, often based on machine learning and “big data”, has **great potential** to identify, quantify, and reduce risks...

But their use can also create **new types of risk**.
Automated Algorithms and Risk
Part 1: The Promise

Better Models, Better Decisions
Early Warning for Critical Events
Identifying Emerging Trends and Patterns
Better models, better decisions

Better **predictive models** can lead to improved organizational decisions, allocation of public resources, and quantification of risks.

Which geographic areas within a city are likely to become crime hot spots?

**Short time horizon:** Targeted police patrols in these areas reduce crime.

**Longer time horizon:** Quantify risk of loss from theft, assault. Also informs city planning and policy.
Early Warning for Critical Events

**Early detection** can reduce costs to society by enabling a targeted and effective response.

**Advance prediction** can both quantify and reduce risk.

- Disease pandemic
- Natural disaster
- Terrorist attack
- Civil unrest

**Twitter Event Surveillance**

Can accurately predict civil unrest up to 1 wk. in advance

Enables earlier detection of emerging disease outbreaks

Can identify emerging human rights issues

(Chen and Neill, KDD 2014)
Identifying Emerging Trends and Patterns

- Trends of opioid use, abuse, addiction, and overdose
- Patterns of patient care that impact health outcomes
- Line-of-duty injuries among uniformed service workers
- Building/neighborhood issues and chronic health conditions
Automated Algorithms and Risk
Part 2: The Perils

Who is responsible when these algorithms fail, particularly when they fail **systematically**?

One major risk area: **fairness**, **bias**, and **discrimination** in algorithmic decision-making
Northpointe’s COMPAS software has been used for criminal justice in many jurisdictions to predict individuals’ re-offending risk.

ProPublica compared COMPAS predictions to observed re-arrests and concluded that COMPAS is racially biased.

Huge potential impacts: civil and criminal liability, loss of reputation, loss of future business, erosion of public trust in civil institutions…
Mitigating risks of algorithmic bias

Need for increased transparency:
- Is the model specification reasonable?
- Is the training dataset representative?
- Is the target variable biased?
- Possible unintended consequences?

Should assess by multiple approaches, both human and automated.

We have developed a novel approach ("bias scan") to audit black-box risk prediction algorithms for fairness, and to correct systematic biases.

When applied to COMPAS, our system revealed 3 biases:
- Number of prior offenses
- Males under age 25
- Females w/ misdemeanors

(Not racial biases, but...)
Conclusions

Automated machine learning algorithms have great, mostly untapped potential to benefit the insurance and reinsurance fields:

**Underwriting** – advance prediction of crisis events, risk estimation

**Risk mapping/surveillance** – early detection of events, trends, patterns

**Loss mitigation** – reducing impact through early and targeted response

But they also create new risks of systematic failures and unforeseen consequences, including the potential for algorithmic bias and discrimination.

These new risks should be carefully considered and mitigated, ideally through a combination of human expert oversight and algorithmic auditing approaches.
Thanks for listening!

More details on my web site:
http://www.cs.nyu.edu/~neill

Or e-mail me at:
daniel.neill@nyu.edu