Technology and the Financial System

SUSAN ATHEY

THE ECONOMICS OF TECHNOLOGY PROFESSOR, STANFORD GSB

A Call For Action for Fin Tech Regulators

Create a “best practices” set of guidelines for evaluating machine learning and AI applications

- Update this regularly, tracking changes
- Approach should be nuanced, by application and model type

Articulate an appreciation for:

- Risks of increased transparency of using algorithms
- Demonstrable error rates
- Ability to identify problems in principle, but inability to take action on all
- Benefits as well as risks of automation

Consider the loss function in each application

- Rational cost-benefit analysis

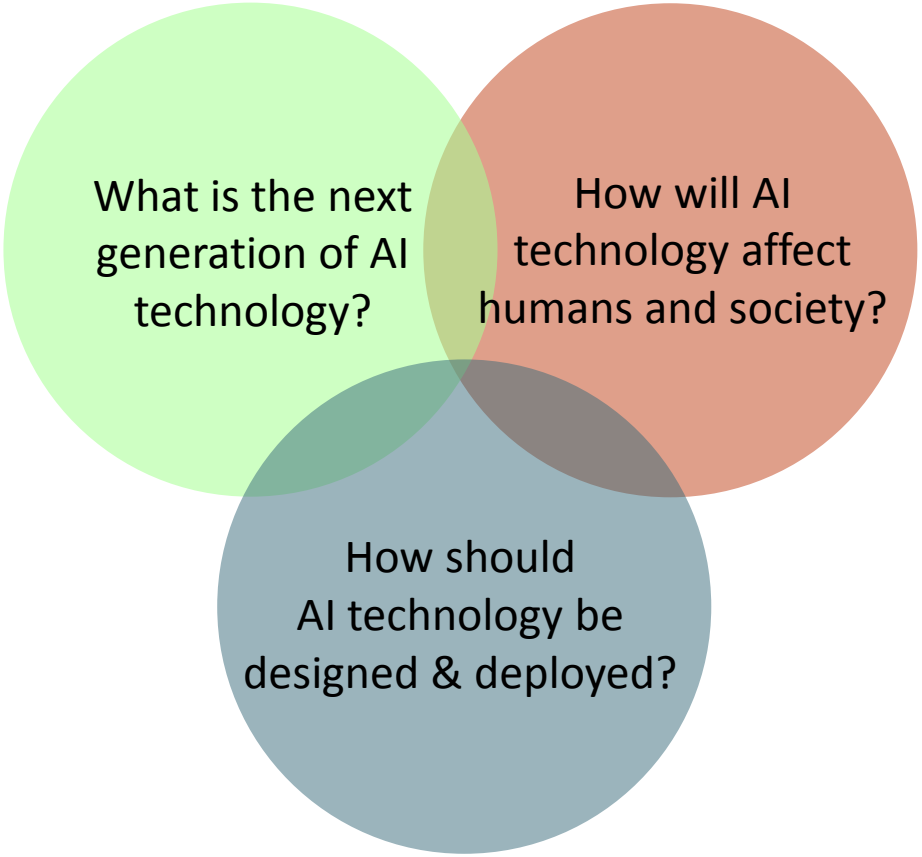
Research agenda

- How to validate and stress test ML models systematically

In age of open source and cloud, consider promoting public good technology and R&D

- Security, identity, etc.

Pressing questions about AI



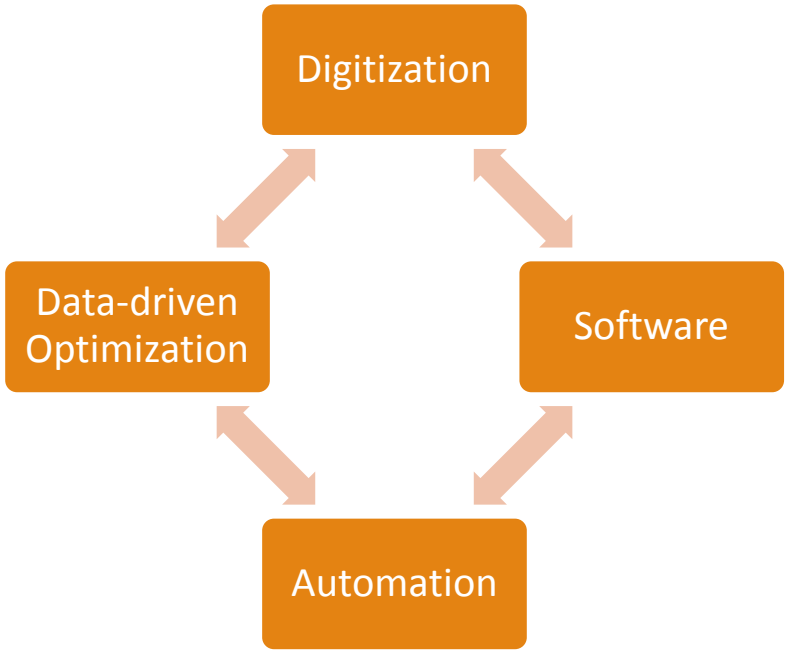
Software is Eating The World
Every Company is a Tech Company

"My own theory is that we are in the middle of a dramatic and broad technological and economic shift in which software companies are poised to take over large swathes of the economy.

More and more major businesses and industries are being run on software and delivered as online services—from movies to agriculture to national defense."

-Marc Andreessen

(2013)



Regulation and Management Closely Related

How does organization manage and optimize implementations of AI?

How do regulators evaluate and guide AI implementations?

Benefits and risks of different implementations

- Human, manual processes
- Simple statistical models
- Supervised ML models
- Active learning/exploration
- Complex AI systems

Production function for a digitized firm

- More decentralized
- Greater use of “black boxes”—even developers don’t understand
- Rules/metrics/processes matter more
- Management as internal regulators

Management/regulation differ for human processes versus algorithms

- Algorithms make mistakes, documented
- Algorithms respond strongly and quickly to incentives
- Need quantitative, quickly measured success metrics

Andrew Ng @AndrewYNg · May 1

If you're trying to understand AI's near-term impact, don't think "sentience." Instead think "automation on steroids."

38 1.1K 1.6K

AI Applications: Automation on Steroids

Many tasks cannot be fully outsourced to black box

- Augment humans
- Augment domain models

Easy to evaluate success quickly

Lots of data

Large feature space

- Hard for human to attend to
- Hard for simple model to capture

Low cost of mistakes

Can improve/retrain faster than the environment changes

Have all relevant scenarios in training data (past or current experiments)

Challenges for Regulation of AI

Challenges with Statistical Approaches

Challenges exacerbated with complex functional forms unless care is taken

- Lack of stability/robustness
- Poor performance when extrapolating
- Correlation v. causation
- Non-obvious omitted variable bias
- Discrimination and fairness

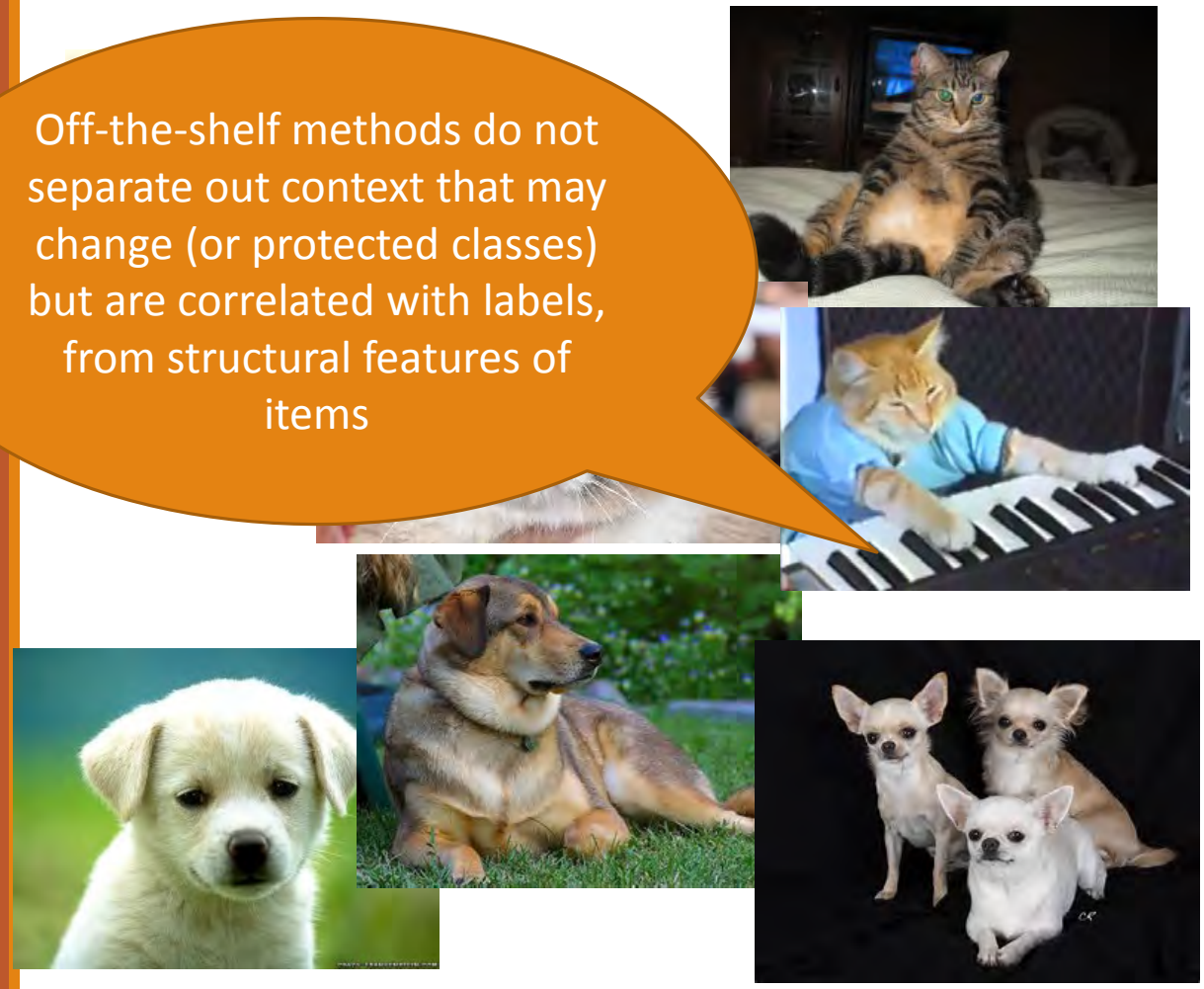
Black box—need analytics to evaluate

- Uncertainty quantification
- Instability/robustness
- Where are biases likely?
- When to trust model v. human?

Machine Learning and AI

Advances in ML dramatically improve quality of image classification

Off-the-shelf methods do not separate out context that may change (or protected classes) but are correlated with labels, from structural features of items



What's Ne
About M

Flexible, rich, data-driven models

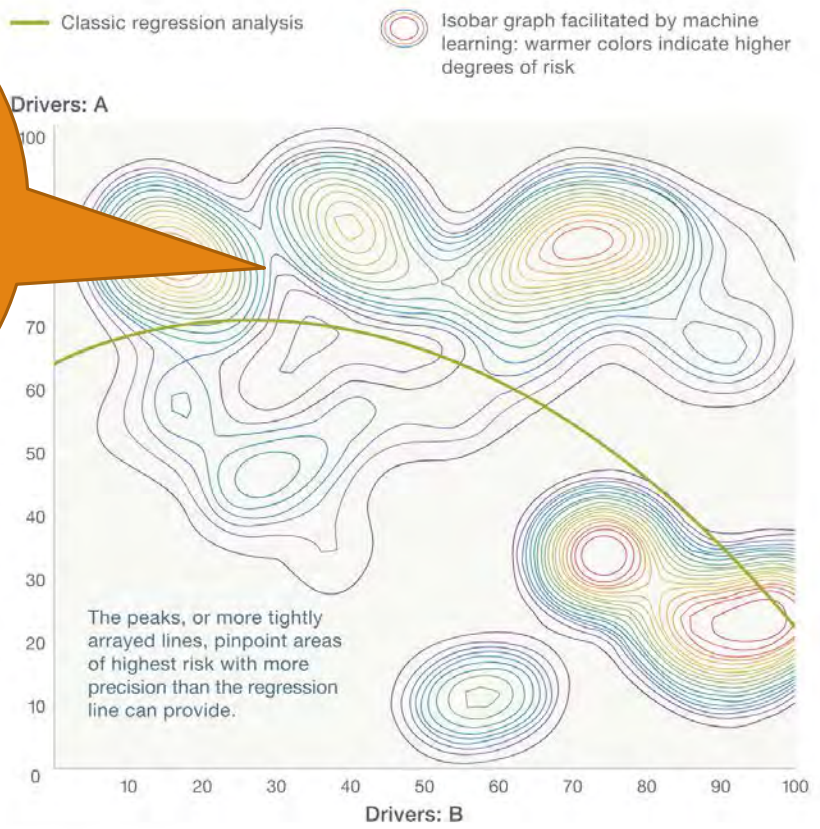
Increase in personalization and precision

Methods to avoid over-fitting

Do we really think this relationship is plausible? Stable, robust, causal?

The contrast between routine statistical analysis and data generated by machine learning can be quite stark.

Value at risk from customer churn, telecom example



McKinsey&Company

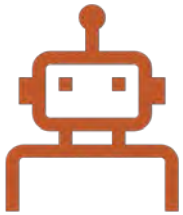
The ML “Production Function”

In industry, ML algorithms are constantly improved

Learning by doing

Randomized experiments with live user traffic

Decentralized innovation



ML algorithms require an objective

Need to decide/agree on goals

Measuring some things easy, others hard

Challenges for Management /Regulation of AI

Challenges with Operationalizing Innovation

Given stability requirements, frequently update model

- Need processes to evaluate systematically, quickly, frequently
- A/B Testing

Multiple objectives

- Some measured well, some poorly
- Some short term, some long term
- End up optimizing for short term, well measured metrics, creating biases and risks

Use surrogates

- Often can be manipulated intentionally or unintentionally
- “Teaching to the test”

Short term metrics: Buzzfeed
What do you get when you optimize for clickthroughs?

15 Of The Creepiest Things Kids Have Ever Said To Babysitters

You'll never look at kids the same way.

Javier Moreno · a half hour ago · 16 responses

How Cat Are You?

Take this quiz MEOW.

Leonora Epstein · an hour ago · 144 responses

37 McDonald's Foods You Probably Haven't Tried

What Your Favorite Breakfast Says About Your Sex Life

One eggs benedickt, please.

Tanya Chen · 4 hours ago · 86 responses

Challenges for Management /Regulation of AI

Challenges with Human-Machine Interaction

Old Approaches

- Manually inspect models, consider implications
- Reason about omitted variables
- Reason about internal, external validity

Systematic Methods for Black Boxes

- People who engineer machines don't have conceptual framework
- Research has not delivered systematic, general-purpose tools that replace old approaches
  - Perhaps because it is hard to do so!
- Even if we could replicate "knowledge" experts would gain from statistical model in old system…
  - How do we communicate these to human users?

How to combine humans and algorithms optimally?

Using Streetview to Predict Safety, Housing Prices

- Get Google Streetview Images
- Extract features for color, texture, shape
- Human labels for safety
- Build a model predicting safety, housing values
- Apply out of sample (other cities, etc.)

http://streetscore.media.mit.edu/about.html

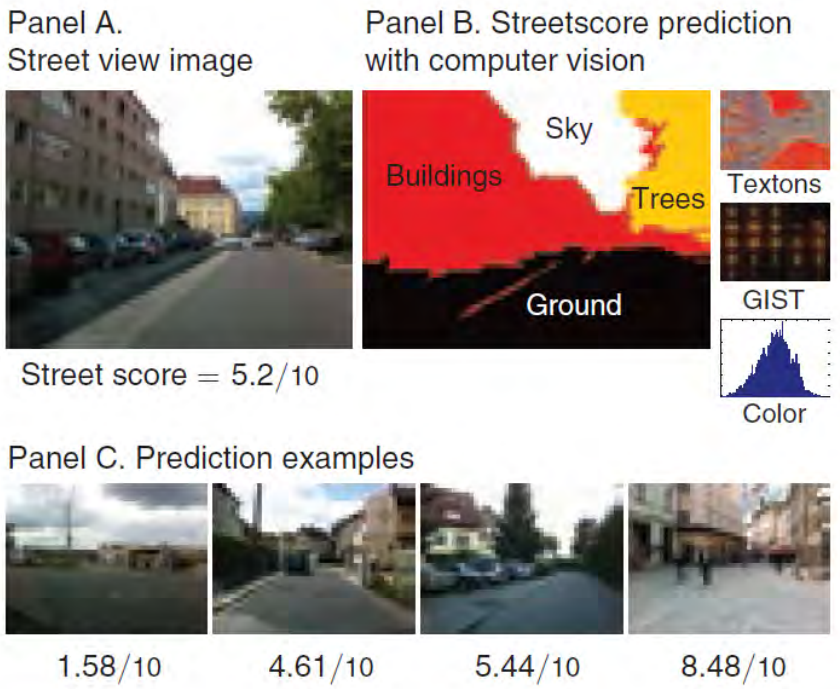


FIGURE 1. COMPUTER VISION TO PREDICT THE PERCEIVED SAFETY OF STREET VIEW IMAGES

Results

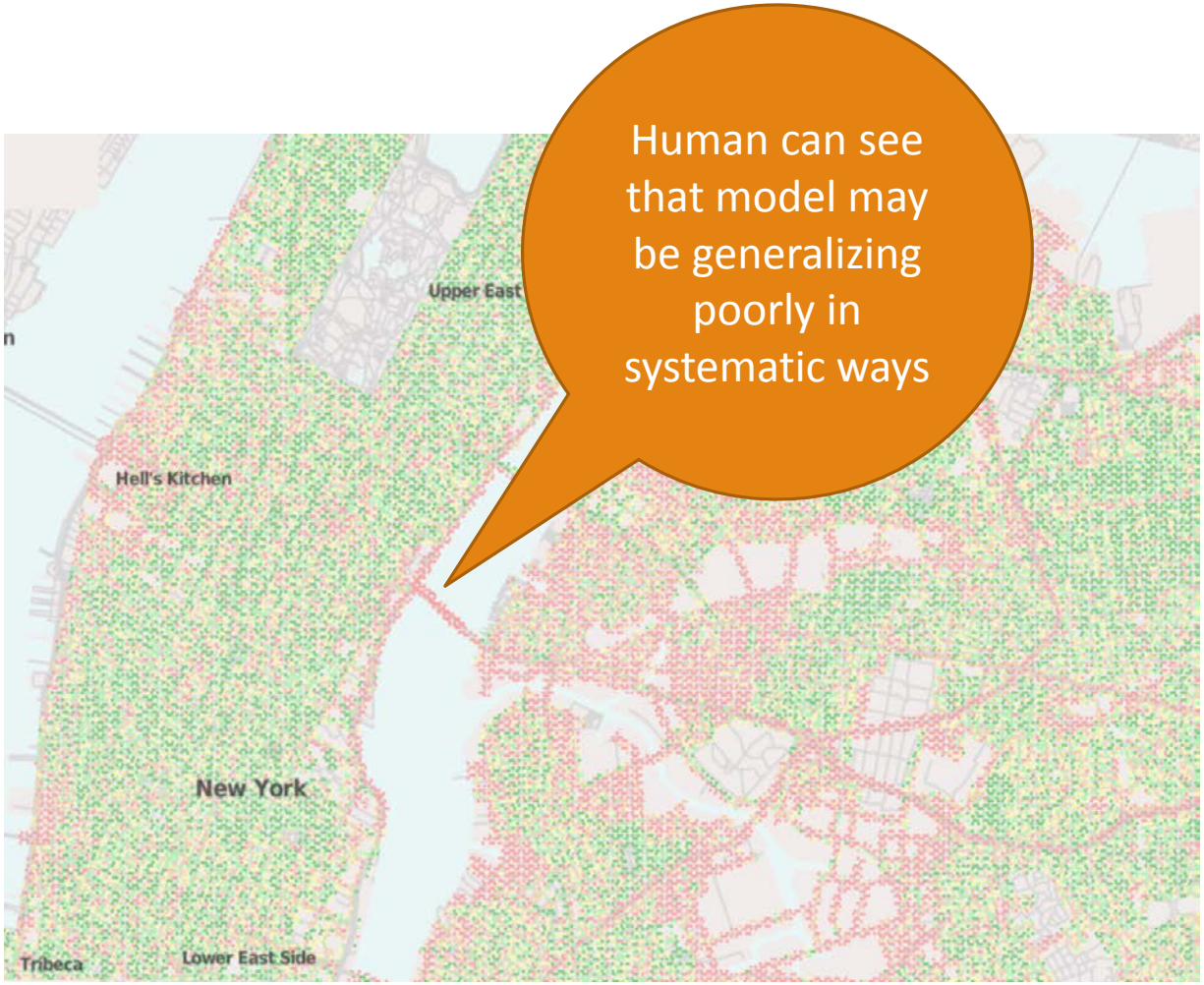
R^2 of 57% in cross-validation samples

Other versions predict land/housing prices

Can then correlate land/housing values to demographics, other variables

Future:

- Track over time
- Relate to policies, other shocks



Challenges for Management /Regulation of AI

Challenges with Using AI in the Wild

Credit Scoring Example

- **Instability** of joint distribution of outcomes, novel features
- **Manipulation** of novel features
- Ever-changing **adverse selection** problem as competing firms change models, marketing strategies

Equilibrium effects

- Agents using ML interact
- Collusion (airline prices)
- Instability (financial market crashes, correlated mistakes across firms)
- Google maps examples

Need models of individual behavior and equilibrium selection to study eqm changes

- Why existing AI R&D is a long way from solving “harder” problems

Measuring Poverty in Sri Lanka with Satellite Imagery

Train predictive model on poverty data

Explain 60% of variation in poverty below 40th percentile

Figure 8: Example Roof Type Classification

Poverty in HD: What Does High Resolution Satellite Imagery Reveal about Economic Welfare?

Ryan Engstrom, Jonathan Hersh, and David Newhouse

http://pubdocs.worldbank.org/en/607414661817437796/Poverty-in-HD-draft-v2-75.pdf

Prediction versus "What-If" Problems

PREDICTION

- Is it a cat?
- Is the review positive?
- Will a user with given characteristics click on the ad?

WHAT-IF

- What will happen to sales if I raise prices?
- What was the ROI on my advertising campaign?
- How will my competitors react if I introduce a new product?

Need designed or "natural" experiments to answer what-if questions
More complex AI systems need to be good at causal inference

Challenges for Management /Regulation of AI

Challenges with bandit exploration, reinforcement learning

Bandits

- Explore/exploit tradeoff
- Learn best of many alternatives
- Factorial: combinations of different choices (email subject line, body, offer)
- Contextual: learn personalized policy
- Need to specify objective, get quick feedback
- Need to be happy with all alternatives

Reinforcement learning

- More dynamics, more choices

Many possible unintended consequences

To learn more about using ML in economics

Some of the resources I have made available

Pitfalls of Pure Prediction

- "Beyond Prediction: Using Big Data for Policy Problems," *Science*, February 3, 2017

Surveys and Overviews

- "The Impact of Machine Learning on Economics," forthcoming in *The Economics of Artificial Intelligence*, NBER volume
- "Machine Learning Methods Economists Should Know About," forthcoming in *Annual Reviews*

Lecture notes on Machine Learning and Causal Inference, Tutorials

- https://www.aeaweb.org/conference/cont-ed/2018-webcasts
- http://bit.ly/2CGLfes includes full class notes from multiple classes, as well as R scripts, tutorials, etc.
- AEA/AFA Lecture on Impact of ML on Economics *https://www.aeaweb.org/webcasts/2019/aea-afa-joint-luncheon-impact-of-machine-learning*

References from my work on ML

Stable/robust prediction and estimation

- "Stable Prediction across Unknown Environments," (with Kun Kuang, Ruoxuan Xiong, Peng Cui, and Bo Li), *Knowledge Discovery and Data Mining*, 2018.
- "Estimating Average Treatment Effects: Supplementary Analyses and Remaining Challenges," (with Guido Imbens, Thai Pham, and Stefan Wager), *American Economic Review*, May 2017
- "A Measure of Robustness to Misspecification" (with Guido Imbens), *American Economic Review*, May 2015, 105 (5), 476-480

Causal inference with Panel Data

- "Matrix Completion Methods for Causal Panel Data Models" (with
- "Synthetic Difference in Differences" (with Dmitry Arkhangelsky, David A. Hirshberg, Guido W. Imbens, Stefan Wager)
- "Ensemble Methods for Causal Effects in Panel Data Settings" (with Mohsen Bayati, Guido Imbens, Zhaonan Qu), *American Economic Review Papers and Proceedings,* 2019

Combining Machine Learning and "Structural Models" of Consumer Behavior in Panel Data

- "Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data," (with David Blei, Robert Donnelly, Francisco Ruiz, and Tobias Schmidt), *American Economic Review Papers and Proceedings*, May, 2018
- "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements," (with Francisco Ruiz and David Blei), forthcoming, *Annals of Applied Statistics*.
- "Counterfactual Inference for Consumer Choice Across Many Product Categories" (with Rob Donnelly, Francisco Ruiz, David Blei)

ML and Causal Inference: Treatment Effects and Assignment Policies

- "Generalized Random Forests," with Julie Tibshirani and Stefan Wager, *Annals of Statistics,* 2019.
- "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests" (with Stefan Wager), *Journal of the American Statistical Association*, 2018.
- "Efficient Policy Learning," with Stefan Wager, 2017.
- "Offline Multi-Action Policy Learning: Generalization and Optimization," (with Zhengyuan Zhou and Stefan Wager)
- "Local Linear Forests," (with Rina Friedberg, Julie Tibshirani, and Stefan Wager), 2018.
- "Recursive Partitioning for Heterogeneous Causal Effects" (with Guido Imbens), *Proceedings of the National Academy of Science* 2016 113 (27) 7353-7360
- "Estimating Treatment Effects with Causal Forests: An Application" (with Stefan Wager)

Contextual Bandits

- "Balanced Linear Contextual Bandits," with Maria Dimakopoulou, Zhengyuan Zhou, and Guido Imbens, *Association for the Advancement of Artificial Intelligence (AAAI),* forthcoming.

Theory and applications of surrogates

- "Estimating Treatment Effects using Multiple Surrogates: The Role of the Surrogate Score and the Surrogate Index" (with Raj Chetty, Guido Imbens and Hyunseung Kang), 2016

A More In-Depth Look…

Managing People v. Algorithms

PEOPLE

- Notice unintended consequences
- Consider how outcomes and processes will be evaluated both subjectively and objectively
- Make tradeoffs

- Mistakes – overweight subjective issues
- Slow, ignore lots of info

ALGORITHMS

- Do exactly what they are told to do
- Will continue to optimize for chosen objectives
- Need to get feedback—so usually guided by short term measures of success
- Fast and efficient
- Art of managing algorithms: defining objectives carefully, considering unintended consequences

Algorithms exhaustively explore and optimize to meet your objective.
If you neglected to mention not to drive off a cliff, you may drive off a cliff very quickly!

What is the Loss Function?

NEED TO BE RIGHT ON AVERAGE	EACH DECISION MUST BE RIGHT
Search engines	Anti-money laundering
Decision support algorithms	Making large loans or deals
Online advertising	

Consider how to combine ML/AI with humans and heuristics to handle high-risk scenarios

Characteristics of Applications Suitable for Machine Learning and AI

ACCURATE, QUICKLY OBSERVED MEASURE OF OUTCOME OF INTEREST

Accurate measure
- Signal to noise ratio: classic issue, requires bigger sample
- Less predictable outcome -> more room for factors that may change over time
- "You get what you pay for": algo optimizes

Quickly observed
- Need to learn and iterate to improve model (can't directly inspect black box)
- Assumptions about stability hard to assess

Loss function: e.g., in financial services
- Errors can be devastating esp. w/ regulators

LARGE DATASET WITH RICH FEATURES

With small number of features, gains to black-box algorithm reduced
- Still helps with non-linearities

Data set must be large to learn complex relationships reliably
- Large relative to noise

More likely to have large dataset when outcome passively collected
- But such outcomes may be noisy measures of final objective

Timing of Observing Outcome

QUICKLY OBSERVED

Classification

- Images
- Character Recognition
- Things that can be evaluated by low-skill human

Short-term loans

Conversion rate of digital offers

LONG TIME LAG

Health interventions

Educational programs

- Retirement savings advice

Long-term loans

Investments in trust and reputation

“Surrogate” Outcomes

EXAMPLES

Health status indicators

Student test scores

Consumer actions on a web site

Changes in credit scores for loan recipients

Clicks on advertisements

Email open rate

PROBLEMS

Link between surrogate metrics and long-term outcomes depend on intervention

- Different drugs have different pathways to success
- Improving test-taking skills versus improving fundamental knowledge

Manipulability

- Send misleading emails
- Click-bait in ads

You get what you pay for!

Short term metrics: Email Campaigns

LONG TERM METRIC: SALES

Reward email marketing team for revenue booked

- Use gold standard A/B testing to evaluate the effectiveness of emails
- Wait several months to improve email copy

Problems

- Too slow for customization, personalization
- Can't test enough alternatives

SHORT TERM METRIC: EMAIL OPEN RATE

Translate open rate to sales using historical data: $/open

What happens to $/open after metric change?

- Declines to 25% of previous level within months

Solutions

- Peer review of email copy
- Focus groups or human judgements
- Delayed bonus for team based on real revenue

Making the Decision versus Mimicking the Decision-Maker

MAKING THE DECISION

Investing in a Company

Outcome=successful investment

Features=characteristics of team, pitch deck, industry, etc.

Problems

- Too long to measure
- Too few observations

MIMICKING DECISION-MAKER

Gather data from previous investment decisions

Estimate how decision-makers behave

- All else equal, less likely to invest in women, or older founders?
- Preference for certain universities?
- Compare across decision-makers

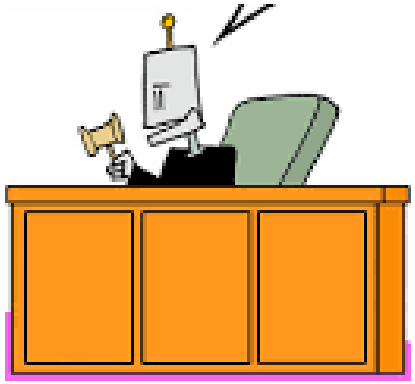
Speed up decision-making, provide decision support

Algorithms versus Humans

Judges decide about releasing suspected criminals on bail

Fit algorithm that mimics judge release rule

- Interpret as what judges would do if they didn't use subjective information
- Finding: fitted rule beats what judges actually do
- Not difficult to preserve racial composition etc. achieved by judges

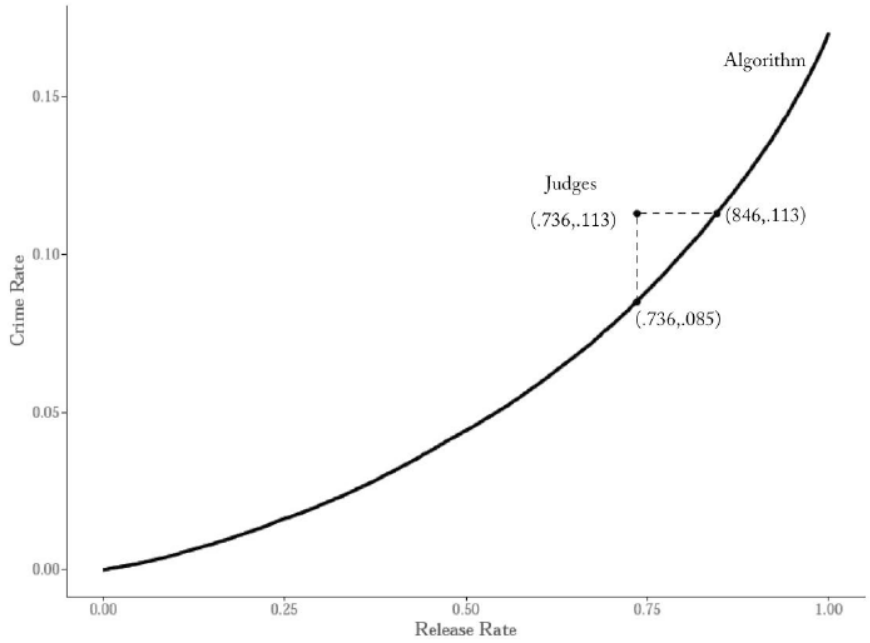


Humans over-react to perceived subjective factors

- Similar evidence from hiring
- Avoid unproductive biases

Kleinberg et al, 2017

Pre-Trial Release Decisions Fit Algorithm to Predict Crime Rate



Application: Resume Screening

First Stage of Hiring
- Lots of data
- Goal 1: mimic human decision-makers for whether person gets interviewed
- Goal 2: predict whether person will be offered a job after gathering more information

Issues
- Reinforcing biases
  - When used for Goal 2, can overcome human biases and be more fair
  - Can intentionally interview a random sample to learn more

Common application
- Screen people for further information gathering
- Works well when humans don't have time to review all relevant info

Application: Monitoring and Incentives

Marketplaces need to provide incentives and screen for quality

- Ratings are noisy, often missing and biased, uncomfortable and time consuming for customers
- Alternative: direct monitoring and feedback to sellers

Approaches

- Gather data passively
- Gather customer satisfaction data from a sample, or passively from customer behavior
- Train a model to estimate quality of service
- Provide feedback and coaching to seller, require training, explicit incentives

Case Study: Risk of Churn v. Prioritization

PREDICTION: WHICH CUSTOMERS AT RISK?

ML Approach

- Features and behaviors leading up to churn
- Static characteristics of high-churn customers
- Output: each customer has a risk score
- Usage: outreach to high-risk customers
- Validation: look at a held-out test set
- Helps also with forecasting and planning

Problems

- Highest-churn customers not the ones with greatest benefit to intervention

OPTIMAL TREATMENT ASSIGNMENT POLICY

Treatment Effect Approach

- Need data from different customers receiving different interventions—may need to run a pilot experiment to generate data
- Estimate personalized treatment effects
- Advanced version: online, adaptive experimentation
- Validation: gold standard requires implementing different algorithms for different customer groups and observing churn rate over a long time period

Improvement over Risk-based Approach

- Columbia Study: only 50% overlap between high benefit to intervention v. high risk

Case Study: Targeting Clicks v. ROI in Ads

PREDICTION: WHICH CUSTOMERS WILL CLICK?

ML Approach

- Prediction model relating user features and behavior to click probability
- Facebook has tool to "learn" which consumers will click on your ad and to customize ad copy
- Helps with budgeting and planning

Problems

- Consumers who clicked might have purchased anyway
- Clicky consumers may not buy
- Clicky consumers may be more expensive

OPTIMAL TREATMENT ASSIGNMENT POLICY

Treatment Effect Approach

- Need data from different customers receiving different ads, or no ads—may need to run a pilot experiment to generate data
- Estimate personalized treatment effects
- Advanced version: online, adaptive experimentation
- Validation: gold standard requires implementing different algorithms for different customer groups and observing purchase rate over a long time period

Improvement over Risk-based Approach

- Can be dramatically different—opposite signs even
- But—click prediction much easier than treatment effect estimation
- May not be able to get statistical significance