PART IV.
High-level awareness o f broader landscape in causal reasoning

Outline

- Discovery of causal relationships from data
- Heterogeneous treatment effects
- Machine learning, representations and causal inference
- Reinforcement learning and causal inference
- “Automated” causal inference

Causal discovery

Effects of causes and causes of effects

- We discussed causal inference: effects of causes
- But a complementary question is causal discovery
  - [Local] Causes of effects
  - [Global] Mapping out causal mechanisms
- In general, a harder problem.
- See Causation [Spirtes (2000)] and Elements of Causal Inference (Scholkopf et al. 2017).

Heterogenous treatment effects

Average causal effect does not capture individual-level variations

- Stratification is one of the simplest methods for heterogenous treatment by strata
- Typical strata are demographics.
- Need more data to statistically detect differences

- For high-dimensions, can use machine learning methods like random forests [Athey and Wager, 2015]

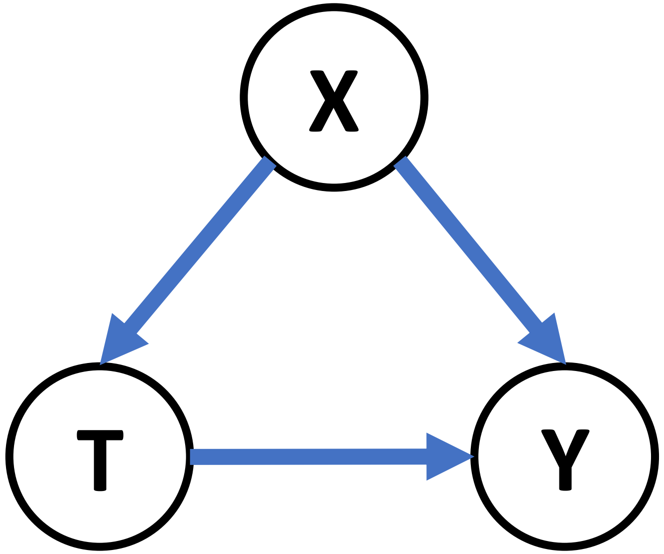
Machine learning and causal inference

Causal inference as a (counterfactual) prediction problem

Causal inference ⇔ robust prediction

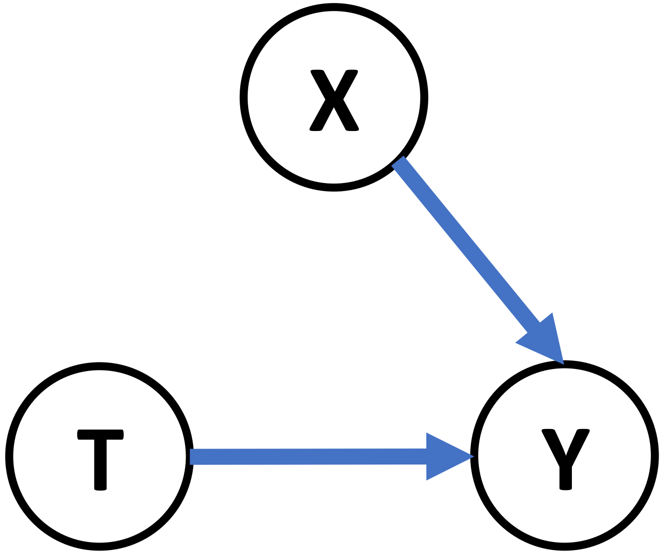
(Supervised) ML	Causal inference
Predicted value under the training distribution *P(X,y)*.	Predicted value under the counterfactual distribution *P′(X,y)*.
$\boldsymbol{P}(\boldsymbol{X}, \boldsymbol{y})$: $y = k(X) + \epsilon$	$\boldsymbol{P}'(\boldsymbol{X}, \boldsymbol{y})$: $y = ?$

Causal inference: A special kind of domain adaptation



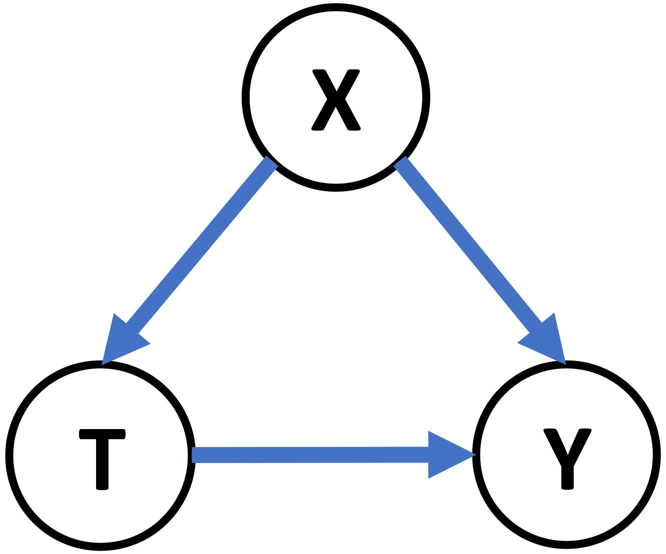
P(Y,T,X)

Observed data



P*(Y,T,X)

Randomized experiment



P**(Y,T,X)

Another domain

Predicting the counterfactual ⇔ Causal Inference

Predicting Individual treatment effects can be considered as domain adaptation

--Use regularization and transformation of input features [Johansson 2016]

Generalizing prediction to new domains

-- Selection bias or covariate shift [Barenboim and Pearl 2013]

-- If predictive model generalizes to new domains, can be considered "causal" [Peters et al. 2015]

Causal inference and machine learning

Machine learning

Use causal inference methods for robust, generalizable prediction.

Causal inference

Use ML algorithms to better model the non-linear effect of confounders, or find low-dimensional representations.

In general, be wary of methods that have not been empirically tested, especially ones that you do not understand.

Reinforcement learning and causal inference

Generalizing a randomized experiment

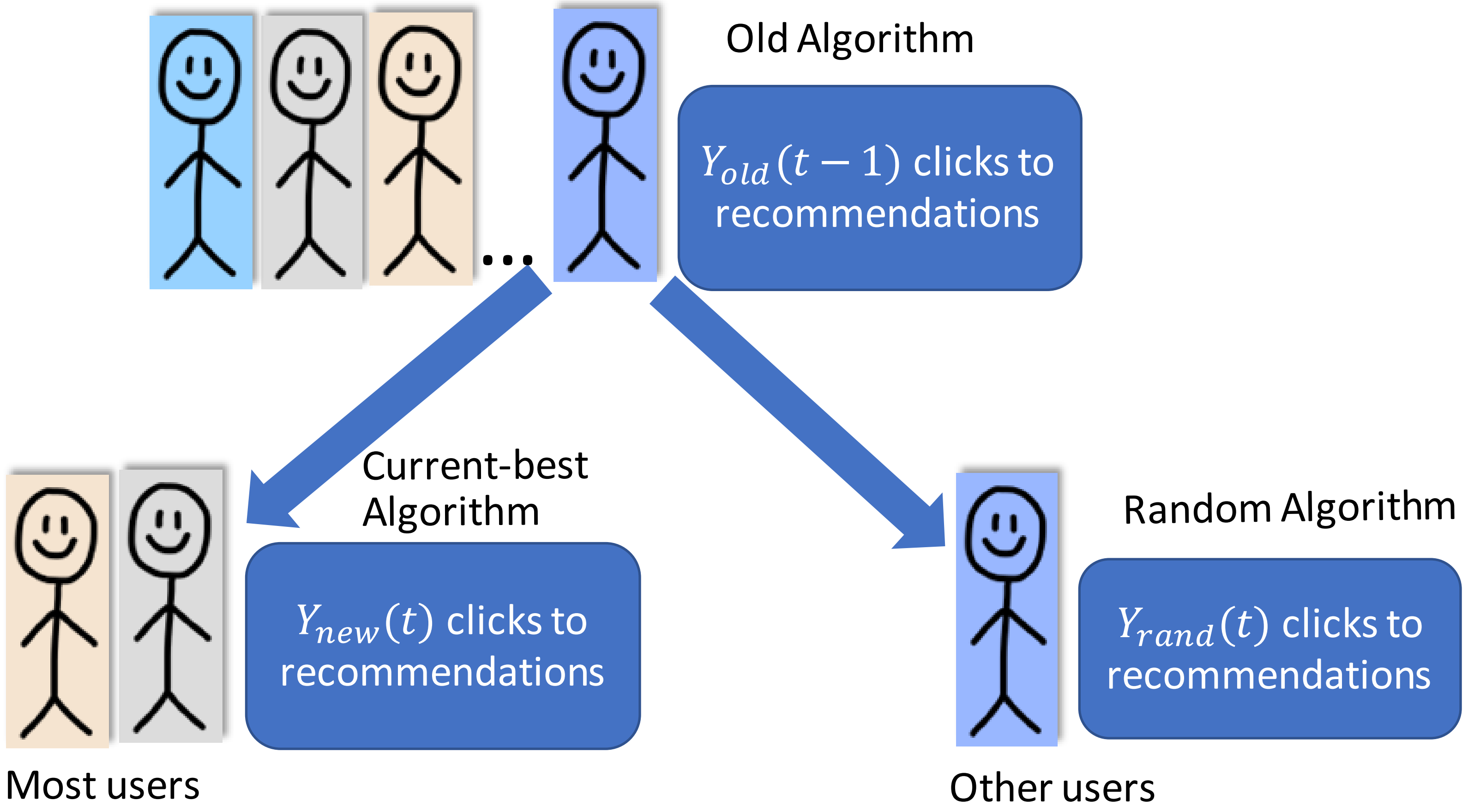
A/B test → Multi-Armed Bandits → Markov Decision Processes → POMDPs

Efficient randomized experiment: Multi-armed bandits

Two goals:

1. Show the best known algorithm to most users.
2. Keep randomizing to update knowledge about competing algorithms.

“Explore and Exploit” strategy



Algorithm: ε-greedy multi-armed bandits

Repeat:

> **(Explore)** With low probability ε, choose an output item randomly.
>
> **(Exploit)** Otherwise, show the current-best algorithm.

Use CTR results for Random output items to train new algorithms offline.

Practical Example: Contextual bandits on Yahoo! News

Actions: Different news articles to display

A/B tests using all articles inefficient.

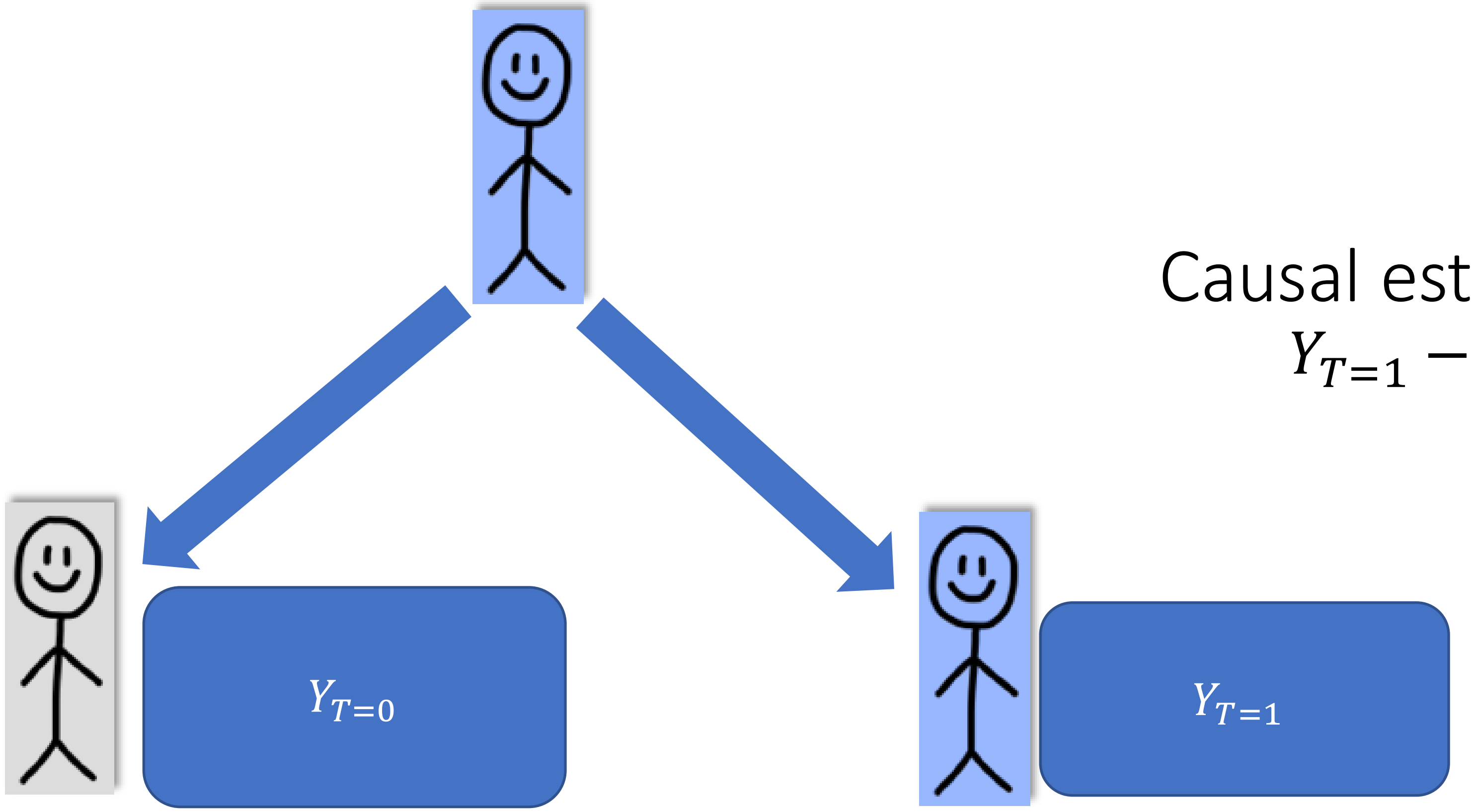
Randomize the articles shown using ε-greedy policy.

Better: Use context of visit (user, browser, time, etc.) to have different current-best algorithms for different contexts.

Li-Chu-Langford-Schapire (2010)

Many of these techniques can be combined

Remember, we are always looking for the ideal experiment with multiple worlds



Cloned user

Causal estimate

$$Y_{T=1} - Y_{T=0}$$

Example: Randomization + Instrumental Variable

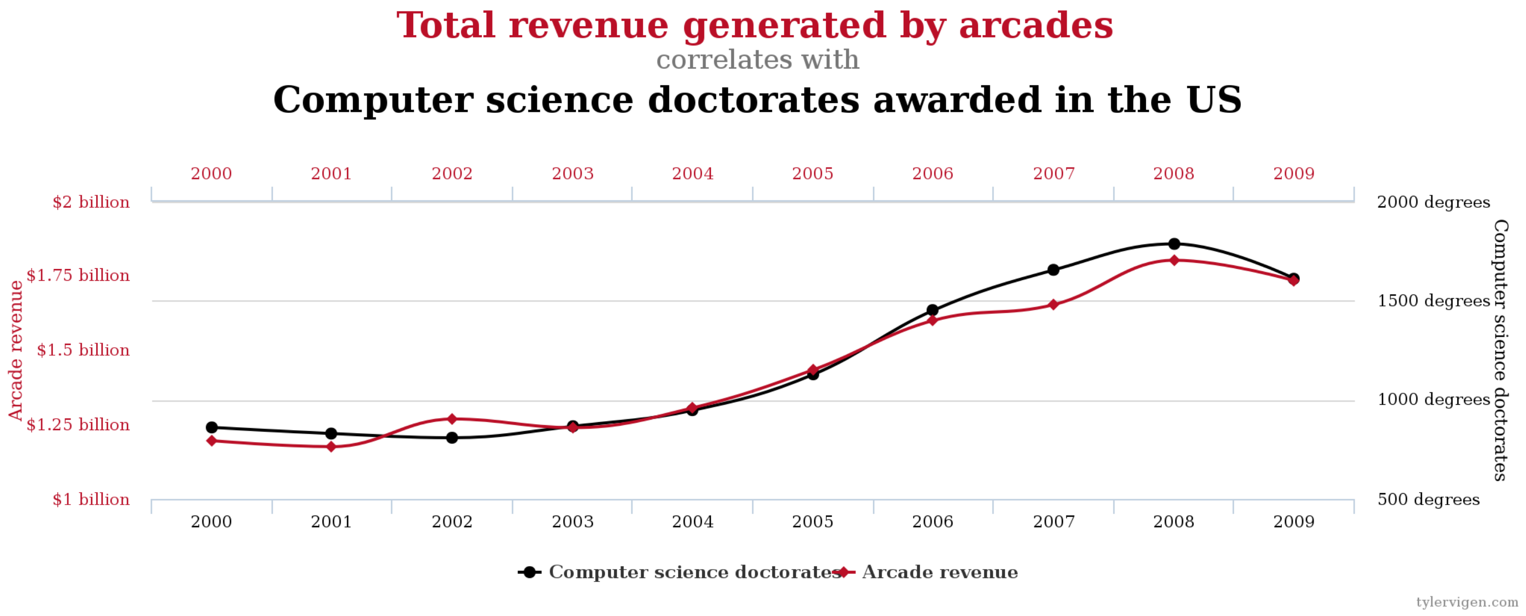
Treatment example: You cannot randomize who exercises, but maybe can provide incentives to join the gym.

Algorithm example: You cannot remove recommendations at random, but could advertise a focal product to a random subset of people on the homepage.

Conclusions

Causal inference is tricky

Correlations are seldom enough. And sometimes horribly misleading.



Always be skeptical of causal claims from ~~observational~~ any data.

More data does not automatically lead to better causal estimates.

http://tylervigen.com/spurious-correlations

Causal inference: Best practices

Always follow the four steps: *Model, Identify, Estimate, Refute.*

--Refute is the most important step.

Aim for simplicity.

--If your analysis is too complicated, it is most likely wrong.

Try at least two methods with different assumptions.

--Higher confidence in estimate if both methods agree.

Thank you!

Emre Kiciman, Amit Sharma (Microsoft)

@emrek, @amt_shrma

Tutorial and other resources will be posted at:
http://causalinference.gitlab.io

DoWhy library can be accessed at
http://causalinference.gitlab.io/dowhy