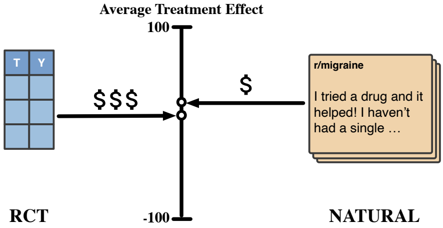
NATURAL

End-To-End Causal Effect Estimation from Unstructured Natural Language Data

Nikita Dhawan[1,2]
Leonardo Cotta[2]
Karen Ullrich[3]
Rahul G. Krishnan[1,2]
Chris J. Maddison[1,2]



[1]University of Toronto [2]Vector Institute [3]Meta AI

Cause and effect questions are everywhere, and often critical.

Repurposing Experiences

Cause and effect questions are everywhere, and often critical.

We don't have the luxury of rigorously testing every possible answer, *e.g.* with a randomized experiment.

Repurposing Experiences

Cause and effect questions are everywhere, and often critical.

We don't have the luxury of rigorously testing every possible answer, *e.g.* with a randomized experiment.

We must prioritize.

Repurposing Experiences

Cause and effect questions are everywhere, and often critical.

We don't have the luxury of rigorously testing every possible answer, *e.g.* with a randomized experiment.

We must prioritize.

How do we choose which potential answer gets millions of dollars to be tested?

Repurposing Experiences

Cause and effect questions are everywhere, and often critical.

We don't have the luxury of rigorously testing every possible answer, *e.g.* with a randomized experiment.

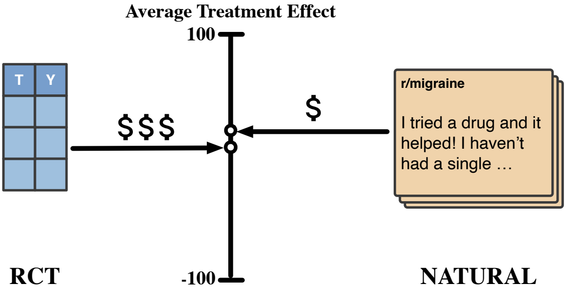
We must prioritize.

How do we choose which potential answer gets millions of dollars to be tested?

Can we learn from existing experiences?

Outline

What is NATURAL?



- A treatment effect estimation pipeline,
- from unstructured natural language data to average treatment effects (ATE),
- built with large language models (LLMs).

Outline

Inverse Propensity score Weighting (IPW)

A standard estimator under classical causal assumptions:

$$\tau_{\mathrm{IPW}} = \mathbb{E}_{X,T,Y}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right]$$

Inverse Propensity score Weighting (IPW)

A standard estimator under classical causal assumptions:

$$\tau_{\text{IPW}} = \mathbb{E}_{X,T,Y}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right]$$

where,

treatments: $T \in \{0, 1\}$,

potential outcomes: $Y(1), Y(0) \in \{0, 1\}$,

observed outcomes: $Y = TY(1) + (1-T)Y(0)$,

covariates or confounders: X,

propensity score: $e(x) = P(T = 1 \mid X = x)$.

Law of total expectation:

$$\tau = \mathbb{E}_{X,T,Y}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right] = \mathbb{E}_{R}\left[\mathbb{E}_{X,T,Y|R}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right]\right],$$

where R denotes unstructured natural language reports.

NATURAL: A TEXT-CONDITIONED ESTIMATOR

Law of total expectation:

$$\tau = \mathbb{E}_{X,T,Y}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right] = \mathbb{E}_{R}\left[\mathbb{E}_{X,T,Y|R}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right]\right],$$

where R denotes unstructured natural language reports.

A Monte Carlo estimate over reports:

$$\hat{\tau}_{\text{NATURAL}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{x,t,y} P(X=x, T=t, Y=y|R_i)\left[\frac{ty}{\hat{e}(x)} - \frac{(1-t)y}{1-\hat{e}(x)}\right].$$

NATURAL: A TEXT-CONDITIONED ESTIMATOR

Law of total expectation:

$$\tau = \mathbb{E}_{X,T,Y}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right] = \mathbb{E}_{R}\left[\mathbb{E}_{X,T,Y|R}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right]\right],$$

where R denotes unstructured natural language reports.

A Monte Carlo estimate over reports:

$$\hat{\tau}_{\text{NATURAL}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{x,t,y}\underbrace{P(X=x,T=t,Y=y|R_i)}_{\text{LLM-estimated conditionals}}\left[\frac{ty}{\hat{e}(x)} - \frac{(1-t)y}{1-\hat{e}(x)}\right].$$

NATURAL: A TEXT-CONDITIONED ESTIMATOR

Law of total expectation:

$$\tau = \mathbb{E}_{X,T,Y}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right] = \mathbb{E}_{R}\left[\mathbb{E}_{X,T,Y|R}\left[\frac{TY}{e(X)} - \frac{(1-T)Y}{1-e(X)}\right]\right],$$

where R denotes unstructured natural language reports.

A Monte Carlo estimate over reports:

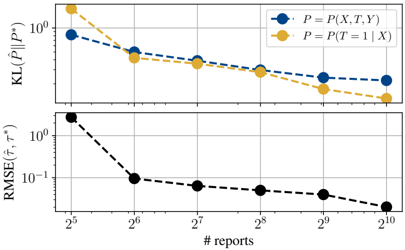
$$\hat{\tau}_{\text{NATURAL}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{x,t,y}\underbrace{P(X=x, T=t, Y=y|R_i)}_{\text{LLM-estimated conditionals}}\left[\frac{ty}{\hat{e}(x)} - \frac{(1-t)y}{1-\hat{e}(x)}\right].$$

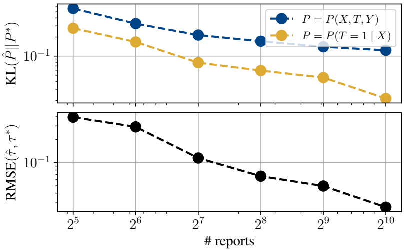
See our paper for different variants of NATURAL!

Outline

How well does NATURAL estimate observational distributions from self-reported data?

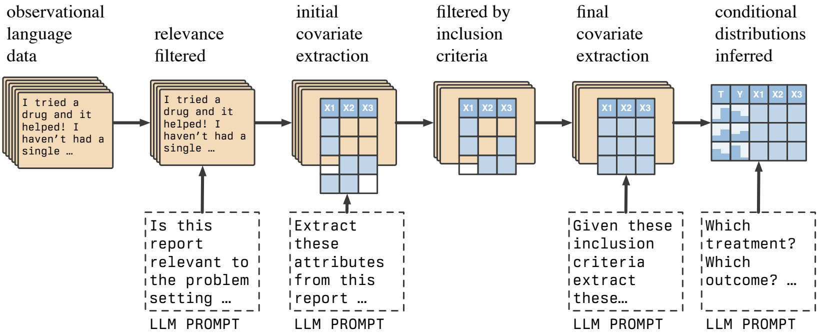
How well does NATURAL estimate observational distributions from self-reported data?





For Hillstrom (left) and Retail Hero (right), the KL divergence between estimated joint and propensity distributions and their true counterparts reduces with increasing number of reports (top), as does the RMSE between the NATURAL estimate and true ATE (bottom).

Filtering raw data



Collect and filter reports relevant to the setting.

Filter reports by inclusion criteria.

Extract (X, T, Y), conditional on text.

How does NATURAL perform in the real world?

How does NATURAL perform in the real world?

We constructed four real-world clinical datasets and compared NATURAL estimates to corresponding randomized controlled trials.

	Tuned	Held-out		
	Semaglutide vs. Tirzepatide (weight loss $\geq$ 5%)	**Semaglutide vs. Liraglutide (weight loss $\geq$ 10%)**	**Erenumab vs. Topiramate (% discontinued)**	**OnabotulinumtoxinA vs. Topiramate (% discontinued)**
	NCT03987919	NCT03191396	NCT03828539	NCT02191579
Treatment effect in real-world RCT	10.11	−14.70	28.30	41.00
NATURAL using social media data	**9.06**	**−16.57**	**29.05**	**42.53**

NATURAL predictions fall within three percentage points of clinical trial ATEs.

Outline

Future work

- How robust is NATURAL to the assumptions it relies on?

Future work

- How robust is NATURAL to the assumptions it relies on?
- Can we combine multiple data sources for effect estimates?

Future work

- How robust is NATURAL to the assumptions it relies on?
- Can we combine multiple data sources for effect estimates?
- How does NATURAL perform at larger scales and in diverse settings (*e.g.* social sciences)?

Thank you!

Leonardo Cotta

Karen Ullrich

Rahul G. Krishnan

Chris J. Maddison

University of Toronto

Vector Institute

Meta