Get to the Bottom: Causal Analysis for User Modeling

Shi Zong[*], Branislav Kveton[#], Shlomo Berkovsky [†]
Azin Ashkan[▽], Zheng Wen[#]
[*]The Ohio State University, [#]Adobe Research, [†]CSIRO, [▽]Google Inc.

Motivation

- Weather affects our mood and behavior, and through them many aspects of our lives

When it is sunny, people become happier and smile

When it rains, people get depressed

- We present the first causal analysis of *how weather affects TV content consumption* patterns

Application

- Causal analysis results can be used for better adaptive *personalized recommendations* of TV and video

Causality

- Setup:
  - *Units* $i \in [n]$ are TV viewing events
  - *Treatment* at event i, $T_i \in \{0, 1\}$, is an indicator of target weather (eg, precipitation) at i
  - *Potential outcomes* at i, $(Y_i(0), Y_i(1))$, are indicators of watched content (eg, Drama programs) in treatment and control
- Goal: Estimate average treatment effect on treated (*ATT*)

$$\mathrm{ATT} = \mathbb{E}_{i:T_i=1}[Y_i(1)] - \mathbb{E}_{i:T_i=1}[Y_i(0)]$$

Treatment group | Treatment outcome | Control outcome

- Problem: Outcome of not treating on treated, $\mathbb{E}_{i:T_i=1}[Y_i(0)]$, is not observed ($\mathbb{E}_{i:T_i=0}[Y_i(0)]$ does not approximate it because the treatments are not assigned at random)
- Key challenge: Balance the distributions of treated and control events
- Solution: *Nearest neighbor matching* on the covariates of events, and then

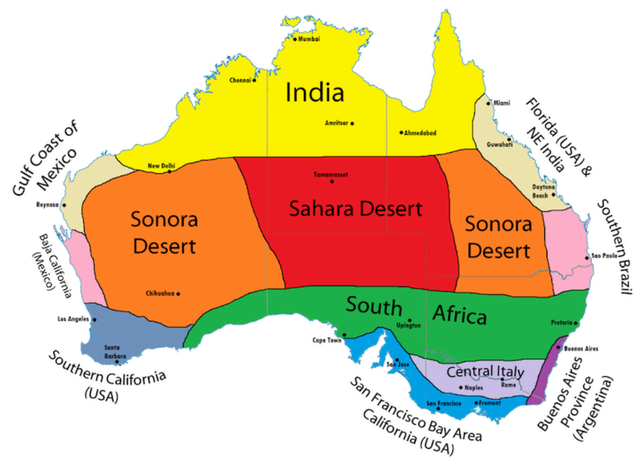
$$\mathrm{ATT} \approx \frac{1}{n_T} \sum_{i:T_i=1} (Y_i(1) - Y_{\pi(i)}(0))$$

Matched control event to treated event i

- *Covariate* is the profile of the user and the time at i

Dataset

- Australia-wide dataset of 10M catch-up TV viewing events from February to September 2012
  - More than 0.6M users all over Australia
  - More than 11K unique programs

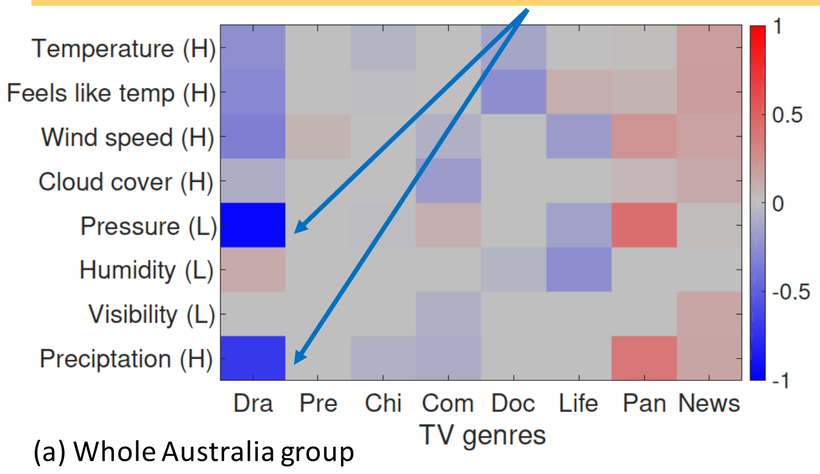


Category	Frequency	Category	Frequency
Drama	19.51%	Pre-school	19.31%
Children	17.01%	Comedy	11.37%
Docs	10.61%	Lifestyle	8.06%
Panel	5.95%	News	4.10%
Arts	2.69%	Education	0.58%
Kids	0.50%	Sport	0.24%
Indigenous	0.05%	Shop	0.02%

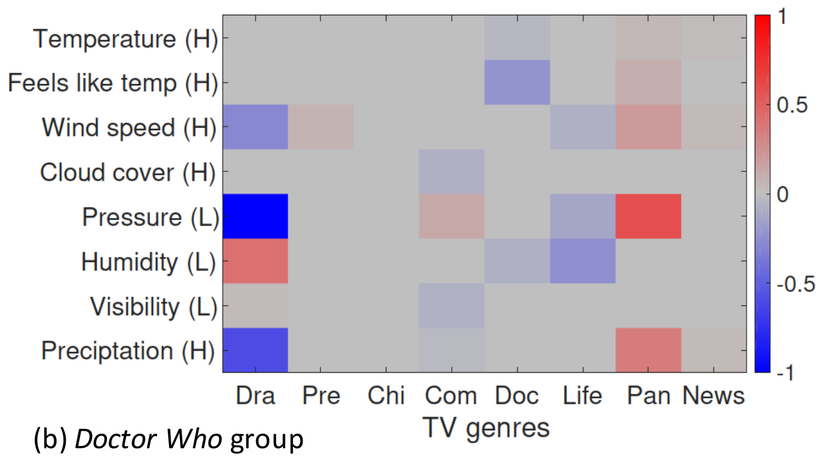
Weather attribute	Treated
Temperature	High
Feels-like temperature	High
Wind speed	High
Cloud cover	High
Pressure	Low
Humidity	Low
Visibility	Low
Precipitation	High

Empirical Results

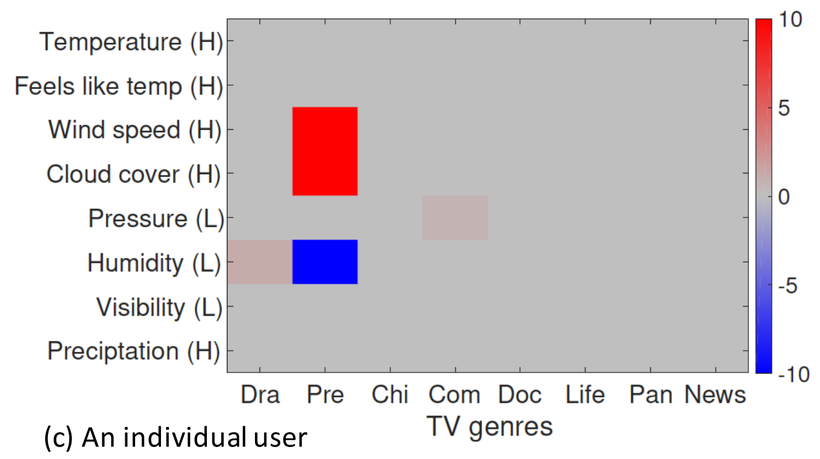
Example: When pressure and precipitation is low (rain), the frequency of watching Dramas drops significantly. Possible explanation: children cannot play outside and watch TV at home



(a) Whole Australia group



(b) *Doctor Who* group



(c) An individual user

Significant changes in ATTs of all weather treatments on 8 most popular genres for different treated groups.

Significant increase in red, significant decreases in blue, insignificant effects in gray.

There is a tradeoff between the size of treated group and the significances of effects.

- Larger groups tend to have sufficient number of treated units, but results can't be used for personalization.
- Only limited individuals have enough treated units for getting significant causal effects.

Our causal findings confirms that there are causal relations between weather and users' TV watching patterns