Optimal Treatment Assignment to Evaluate Demand Response

Baosen Zhang
Electrical Engineering
University of Washington
Grid Science Winter School & Conference
Jan 12, 2017

Joint work with
Pan Li, Yize Chen
Introduction

• Demand Response (DR): send a signal to elicit a change in customer demand
• Change in price, text message, etc.
Introduction

Standard setup for demand response (DR):
1. Direct load control
2. Indirect control:
   – Each user has some utility function (public or private)
   – Maximize the social welfare

Our setting: no direct control and no detailed information

This talk:

How to learn the impact of demand response
Problem Setup

Stylized setup:

• Utility sends a signal, 0 or 1, to a user
  – 1: perform demand response
  – 0: do nothing (or no signal to the user)

• Quantity of interest: causal impact of DR

\[ \text{Consumption|DR} - \text{Consumption|No DR} \]
Challenges

No counterfactual observation

High Dimensionality of Covariates

- Temperature
- Type of day
- Size of house
- EVs
- PVs
- Appliances
- Interaction terms

Most of time a user is not called for DR
- E.g., a user can be called no more than 5 times in one month
Overcoming the Challenges

Estimating an effect under *infrequent signaling* with a *large number of covariates* is a hard problem

- Existing estimation techniques performs poorly

Our approach: *strategically* signaling

- Carefully choose DR signals based on the covariates

Result: We show an optimal estimation strategy with high dimensional covariates
Outline

- Linear model
- Signaling strategy
- Theoretical Analysis
- Simulation with real building data
- Online problem
Additive Linear Model

\[ y_i = \beta x_i + \gamma^T z_i + \varepsilon_i, \]

\( \beta \): the causal impact of DR signal
\( \gamma \): impact of other covariates, vector of dimension \( d \)

Observe \( n \) samples:
\( y_{1}, y_{2}, \ldots, y_{n} \)
\( x_{1}, x_{2}, \ldots, x_{n} \)
\( z_{1}, z_{2}, \ldots, z_{n} \)

Consumption \hspace{1cm} DR signal \hspace{1cm} Covariates \hspace{1cm} Noise

Learn This

\{0,1\}
Estimation Problem

Estimate $\beta$ (impact of DR)

- Given $z_1, z_2, \ldots, z_n$
- Limited signaling: design $x_1, x_2, \ldots, x_n$, at most $k$ of $x_i$ can be 1 ($k \ll n$)
- Observe $y_1, y_2, \ldots, y_n$

$\beta$ : estimate of $\beta$

- Unbiased
- Minimize $Var(\beta)$

High dimensional setting: $d \approx n$

A designer can optimize of the signaling strategy
Standard Practice

• Signals are randomly assigned
  – E.g., $k/n = 1/3$, $x_{\downarrow i} = 1$ with probability $1/3$
• Metric: variance of the estimate, $Var(\beta)$
• High dimension: $d = n - 1$

Every household is treated equally.
Standard Practice

- Signals are randomly assigned
  - E.g., \( k/n = 1/3 \), \( x_i \) with probability \( 1/3 \)
- Metric: variance of the estimate, \( \text{Var}(\beta') \)
- High dimension: \( d = n - 1 \)

Run a linear regression:

\[ y_i = \beta x_i + \gamma^T z_i + \varepsilon_i, \]

Variance does not decrease!
Standard Practice

Method 1: Predict then subtract
• Fit the best predictive model, then subtract out the prediction to find the impact of DR

Estimating $\nu$ is hard!

Method 2: Difference-in-Means
• Ignore covariates, pretend the model is

$$y_{\downarrow i} = \beta x_{\downarrow i} + \epsilon_{\downarrow i}$$

Throwing information away as noise!
Our approach

- Use information in the covariates
- Don’t try to do prediction
- Strategically assign signals

Assigning $DR$ signals allows us to tradeoff between the estimation of the two types of parameters.
Running linear regression, the variance of the estimator of beta is given by

\[ y_i = \beta x_i + \gamma^T z_i + \varepsilon_i, \]

\[ \text{Var} \left( \hat{\beta} \right) = \frac{\sigma^2}{x^T P x} \]

Where

\[ P = I - Z (Z^T Z)^{-1} Z \]

- \( x \): vector of DR signals
- \( Z \): matrix of covariates
Optimization Problem

\[
\begin{align*}
\text{minimize} & \quad \text{Var} \ \hat{\beta} = \frac{\sigma^2}{x^T P x} \\
\text{subject to} & \quad \sum_{i=1}^{n} x_i = k \quad \text{Limited signals} \\
& \quad x_i \in \{1, 0\}.
\end{align*}
\]

\[
\begin{align*}
\text{maximize} & \quad x^T P x \\
\text{subject to} & \quad \sum_{i=1}^{n} x_i = k \\
& \quad x_i \in \{1, 0\}.
\end{align*}
\]

- Non-trivial problem:
  - Non-convex, binary variables
- Is it worth solving? How to solve it?

Random assignment

\[
\begin{array}{c}
50 \quad 100 \quad 150 \quad 200 \\
3.8 \quad 4 \quad 4.2 \quad 4.4
\end{array}
\]
Optimal Assignment

• A lower bound: No strategy can achieve a better reduction in variance than $1/n$

• Two questions:
  – Can we achieve this rate? Yes, there exist an assignment
  – Can we solve the problem efficiently? Yes, relaxation

Look at rate first
Best Rate

\[
\begin{align*}
\text{minimize} \quad & \quad \text{Var} \hat{\beta} = \frac{\sigma^2}{x^T P x} \\
\text{subject to} \quad & \quad \sum_{i=1}^{n} x_i = k \\
& \quad x_i \in \{1, 0\}.
\end{align*}
\]

• Result: There exist a solution such that \( \text{Var}(\beta) \) scales as \( 1/n \), as long as \( d < n \) and \( k/n > \epsilon \), for some fixed \( \epsilon \).

• Contrast: If \( x_i \) are randomly assigned, then \( \text{Var}(\beta) \) stays constant if \( d \) is close to \( n \), for all values of \( k \).
Example

• Synthetic data:
Achieving Optimal Rate

\[ y_i = \beta x_i + \gamma^T z_i + \varepsilon_i, \]

Dimension \( d \)

- Look at the extreme case where \( d = n - 1 \), hardest case to learn \( \beta \)
- Quantity of interest: \( x^T P x \)
- \( P \) is a projection matrix:

\[
P = I - Z(Z^T Z)^{-1} Z^T = y y^T, \quad Z^T y = 0, \quad ||y||_2 = 1
\]

Goal: maximize

\[
(x^T y)^2
\]
Null Space

- Assume $Z$ has random Gaussian entries, is $n$ by $d-1$

  $null(Z^\top)$: has a basis with i.i.d. Gaussian entries

  $y$: normalized version

- Maximize $(x^\top y)^2 = (\sum x_i y_i)^2$

Choose positive values
Extreme Case

Max \((x^T y)^2\)
- The information from each signal is not equal
- Strategically assign to get the maximum information

The optimal algorithm is easy
- Find \(y\)
- Sort: \(y^{(1)} > y^{(2)} > \ldots > y^{(n)}\)
- Assign \(x=1\) to the largest \(k\) elements

Rate is \(n\) as long as \(k/n > \epsilon\)
Example

- Synthetic data: \( d = n-1, \ k/n = 1/3 \)
General Settings

This is actually a graph partition problem:

\[
\begin{align*}
\text{maximize} & \quad x^T P x \\
\text{subject to} & \quad \sum_{i=1}^{n} x_i = k \\
& \quad x_i \in \{1, 0\}.
\end{align*}
\]

There is a SDP relaxation with provable gaps.
Quality of Solution

Gaussian Entries

Uniform Entries
Some Real Data

- A hotel in Seattle, with at most 48 covariates including outside temperature, zonal temperature, heating, appliance, etc…
- Train a regression model based on all the data, then simulate DR
  - We can test the impact of covariate dimensions
Estimation Error

- Fix $n=50$, varying $d$

- Fix $d$, varying $n$

- Trying to conduct some trials
Online Setting

• We have considered the offline problem

\[
\begin{align*}
\text{maximize} & \quad x^T P x \\
\text{subject to} & \quad \sum_{i=1}^{n} x_i = k \\
& \quad x_i \in \{1, 0\}.
\end{align*}
\]

• Online Setting: approximate P in an online fashion

• Some preliminary results
Conclusion

• An optimal treatment assignment strategy in the context of demand response
  – It is possible to learn under unfavorable conditions

• Future work:
  – Online algorithm
  – Other response models
  – Learning and optimizing
SDP Relaxation

There is a randomized algorithm to recover a feasible solution \( x \).

Can show

\[
E[\text{recovered solution}] / \text{SDP} \geq \text{const}
\]
Challenges

Three challenges in estimating the impact of DR:

1. The *counterfactual* is not observed: what would have happened if the opposite was done?

2. There are many other *exogenous factors*