The Power of Persuasion Modeling

Michelangelo D’Agostino
Director of Data Science R&D
mdagostino@civisanalytics.com
@MichelangeloDA

Bill Lattner
Senior Data Scientist
wlattner@civisanalytics.com
@wlattner
The Power of Persuasion Modeling

- Introduction to persuasion modeling
  - response modeling vs. persuasion modeling
  - a note on nomenclature
The Power of Persuasion Modeling

- Introduction to persuasion modeling
  - response modeling vs. persuasion modeling
  - a note on nomenclature
- Persuasion modeling methods
- Evaluating persuasion models
The Power of Persuasion Modeling

- Introduction to persuasion modeling
  - response modeling vs. persuasion modeling
  - a note on nomenclature
- Persuasion modeling methods
- Evaluating persuasion models
- Real-world case studies
  - TV promotional ad effectiveness for the Bravo network
  - persuasion in the 2016 election cycle
  - TV promotional ad effectiveness from observational data
Motivation

Many applications across various domains have a similar form. We want to design and target an intervention to maximize some outcome:

- **Marketing**: maximize the return-on-investment of a particular advertising campaign or offer
- **Website or App**: maximize user engagement or click-through-rate
- **Medicine**: maximize “quality adjusted life years” (QALY’s) through medical interventions
- **Politics**: maximize votes by designing the most persuasive messaging to those on the fence
Response Modeling

One common approach to these problems is to target the people most likely to respond to your campaign, offer, or intervention.
Lookalike Modeling

1. Population
   - Start with a large database of people

2. Customer List
   - And start with a smaller list of client data

3. Match
   - Append additional variables to the client data by matching back to population database

4. Model
   - Find patterns within the client data

5. Score
   - Based on these patterns, give each person in the database a score indicating their likelihood to “look like” the customer list

6. List
   - Make a list of individuals with the highest scores

7. Contact
   - Reach out to the population that “looks like” the original customer list

---

#StrataHadoop
But the key question: How do we know that we’re actually adding incremental sales/users/votes and not just finding the people who would have used us or supported us anyway?
Let’s run a thought experiment to evaluate purchase rates between our targets and non-targets.
Users with a higher predicted purchase score are indeed more likely to respond to the offer than those with lower purchase scores…

- **High Purchase Model Scores**
  - Ad: 3.1%
  - No Ad: 3.0%

- **Low Purchase Model Scores**
  - Ad: 0.7%
  - No Ad: 0.3%
...but the ad has very little incremental effect on those with high scores, who would have purchased at basically the same rate without seeing the ad.
However, the ad does seem to have a high incremental effect among those who weren’t already likely to buy.
However, the ad does seem to have a high incremental effect among those who weren’t already likely to buy.

How do we target the people most likely to respond because of the ad and not just people who were likely to respond anyway?
Persuasion Modeling

- Persuasion modeling can overcome some of these shortcomings with response and lookalike modeling.
- Persuasion modeling starts with a randomized controlled experiment and tries to identify the subsets of people that are most likely to respond to the treatment, offer, or message—not just the people who are most likely to respond anyway.
- If done well, persuasion modeling can beat response and lookalike modeling for driving incremental actions.
It All Starts With an Experiment…

Customers ➔ Control Group ➔ Nothing ➔ Survey or Behavioral Outcome

Customers ➔ Treatment Group I ➔ Promo #1 ➔ ...

Customers ➔ Treatment Group II ➔ Promo #2 ➔ ...

#StrataHadoop
Randomized Controlled Experiments

<table>
<thead>
<tr>
<th>purchased?</th>
<th>promotion?</th>
<th>age</th>
<th>state</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>65</td>
<td>WI</td>
<td>$$</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>43</td>
<td>OH</td>
<td>$</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>44</td>
<td>OH</td>
<td>$$</td>
</tr>
</tbody>
</table>
### Randomized Controlled Experiments

<table>
<thead>
<tr>
<th>purchased?</th>
<th>promotion?</th>
<th>age</th>
<th>state</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>65</td>
<td>WI</td>
<td>$$</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>43</td>
<td>OH</td>
<td>$</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>44</td>
<td>OH</td>
<td>$$</td>
</tr>
</tbody>
</table>

\[ Y_i \text{ our outcome of interest for the } i^{th} \text{ person} \]
### Randomized Controlled Experiments

<table>
<thead>
<tr>
<th>purchased?</th>
<th>promotion?</th>
<th>age</th>
<th>state</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>65</td>
<td>WI</td>
<td>$$</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>43</td>
<td>OH</td>
<td>$</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>44</td>
<td>OH</td>
<td>$$</td>
</tr>
</tbody>
</table>

Our treatment indicator variable, which often takes the values 0 for control and 1 for treatment.
### Randomized Controlled Experiments

<table>
<thead>
<tr>
<th>purchased?</th>
<th>promotion?</th>
<th>age</th>
<th>state</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>65</td>
<td>WI</td>
<td>$$</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>43</td>
<td>OH</td>
<td>$</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>44</td>
<td>OH</td>
<td>$$</td>
</tr>
</tbody>
</table>

Other covariates $\mathbf{x}$ that describe each person in our experiment.
Randomized Controlled Experiments

<table>
<thead>
<tr>
<th>purchased?</th>
<th>promotion?</th>
<th>age</th>
<th>state</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
<td>65</td>
<td>WI</td>
<td>$$</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
<td>43</td>
<td>OH</td>
<td>$</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
<td>44</td>
<td>OH</td>
<td>$$</td>
</tr>
</tbody>
</table>

We can calculate the overall effectiveness of the promotion from this data. We typically call this the ATE (average treatment effect):

$$ATE = \left[ \frac{1}{N_T} \sum_{i \in T} Y_i \right] - \left[ \frac{1}{N_C} \sum_{i \in C} Y_i \right]$$
HTE - ATE’s Evil Extension

- The ATE is useful: it allows us to compare different treatments and promotions for overall effectiveness.
- **BUT**, it is a population average. It is entirely possible to have a negative ATE overall, but for some subpopulations to have a positive treatment effect. In allocating promotional efforts, we would like to identify these heterogeneous treatment effects—groups that benefit more from the treatment than others.
HTE - ATE’s Evil Extension

- The ATE is useful: it allows us to compare different treatments and promotions for overall effectiveness.
- **BUT**, it is a population average. It is entirely possible to have a negative ATE overall, but for some subpopulations to have a positive treatment effect. In allocating promotional efforts, we would like to identify these heterogeneous treatment effects—groups that benefit more from the treatment than others.

- First, some extra notation:

  $Y_i(0)$  outcome for the i-th person if they were in the control group
  
  $Y_i(1)$  outcome for the i-th person if they were in the treatment group
The ATE is useful: it allows us to compare different treatments and promotions for overall effectiveness.

**BUT**, it is a population average. It is entirely possible to have a negative ATE overall, but for some subpopulations to have a positive treatment effect. In allocating promotional efforts, we would like to identify these heterogeneous treatment effects—groups that benefit more from the treatment than others.

First, some extra notation:

**individual-level treatment effect:**

\[ \tau_i = Y_i(1) - Y_i(0) \]
The Rubin Causal Model

<table>
<thead>
<tr>
<th>Y(1)</th>
<th>Y(0)</th>
<th>promotion?</th>
<th>age</th>
<th>state</th>
<th>income</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>?</td>
<td>yes</td>
<td>65</td>
<td>WI</td>
<td>$$</td>
</tr>
<tr>
<td>yes</td>
<td>?</td>
<td>yes</td>
<td>43</td>
<td>OH</td>
<td>$</td>
</tr>
<tr>
<td>?</td>
<td>no</td>
<td>no</td>
<td>44</td>
<td>OH</td>
<td>$$</td>
</tr>
</tbody>
</table>

We only observe the values in blue, but we need both $Y_i(0)$ and $Y_i(1)$ to estimate the treatment effect for each person.

**TL;DR?** It’s a missing data problem, and we can do imputation with a predictive model. The model can learn about what would have happened to a treated person by looking at similar controlled people.
A Note on Terminology

- The literature on this type of modeling is spread across many domains. Keep an eye out for the following:
  - persuasion modeling: political science and politics
  - heterogeneous treatment effects modeling (HTE): economics and social science
  - heterogeneous causal effects: economics and social science
  - uplift modeling or net lift modeling: marketing literature

- Note: as a problem domain, this type of modeling is not very commonly discussed in the machine learning and data science communities. But we think it should be!
Persuasion Modeling Methods
A Very Simple Linear Model

\[ Y \sim T + x_1 + \ldots + x_n + T \times x_1 + \ldots + T \times x_n \]

\[ Y_i(0) = Y_{T=0, X=x_i} \]

\[ Y_i(1) = Y_{T=1, X=x_i} \]

\[ \tau_i = Y_i(1) - Y_i(0) \]
A Very Simple Linear Model

\[ Y \sim T + x_1 + \ldots + x_n + T \times x_1 + \ldots + T \times x_n \]

\[ Y_i(0) = Y_{T=0,X=x_i} \]

\[ Y_i(1) = Y_{T=1,X=x_i} \]

\[ \tau_i = Y_i(1) - Y_i(0) \]
A Very Simple Linear Model

\[ Y \sim T + x_1 + \ldots + x_n + T \ast x_1 + \ldots + T \ast x_n \]

\[ Y_i(0) = Y_{T=0,X=x_i} \]

\[ Y_i(1) = Y_{T=1,X=x_i} \]

\[ \tau_i = Y_i(1) - Y_i(0) \]
A Very Simple Linear Model

Main Effects

\[
Y \sim T + x_1 + \ldots + x_n + T \ast x_1 + \ldots + T \ast x_n
\]

\[
Y_i(0) = Y_{T=0,X=x_i}
\]

\[
Y_i(1) = Y_{T=1,X=x_i}
\]

\[
\tau_i = Y_i(1) - Y_i(0)
\]
A Very Simple Linear Model

\[ Y \sim T + x_1 + \ldots + x_n + T \times x_1 + \ldots + T \times x_n \]

\[ Y_i(0) = Y_{T=0, X=x_i} \]
\[ Y_i(1) = Y_{T=1, X=x_i} \]
\[ \tau_i = Y_i(1) - Y_i(0) \]
A Very Simple Linear Model

\[ Y \sim T + x_1 + \ldots + x_n + T \ast x_1 + \ldots + T \ast x_n \]

\[ Y_i(0) = Y_{T=0, X=x_i} \]

\[ Y_i(1) = Y_{T=1, X=x_i} \]

\[ \tau_i = Y_i(1) - Y_i(0) \]

Estimated outcome if person \( i \) was in the control group
A Very Simple Linear Model

\[ Y \sim T + x_1 + \ldots + x_n + T \times x_1 + \ldots + T \times x_n \]

\[ Y_i(0) = Y_{T=0,X=x_i} \]

\[ Y_i(1) = Y_{T=1,X=x_i} \quad \text{Estimated outcome if person } i \]

\[ \tau_i = Y_i(1) - Y_i(0) \quad \text{was in the treatment group} \]
A Very Simple Linear Model

\[ Y \sim T + x_1 + \ldots + x_n + T \times x_1 + \ldots + T \times x_n \]

\[ Y_i(0) = Y_{T=0, X=x_i} \]
\[ Y_i(1) = Y_{T=1, X=x_i} \]

\[ \tau_i = Y_i(1) - Y_i(0) \]

Estimated treatment effect for person \( i \)
A More Advanced Model: Causal Trees

Athey and Imbens, arxiv:1504.01132v3

- A CART-like or random forest-like algorithm but with an altered split criterion for estimating heterogeneous treatment effects.
A More Advanced Model: Causal Trees

Athey and Imbens, arxiv:1504.01132v3

- A CART-like or random forest-like algorithm but with an altered split criterion for estimating heterogeneous treatment effects
- Choose the split variables and split points from the observable covariates to maximize

$$\frac{1}{N} \sum_{i} \hat{\tau}^2$$
A More Advanced Model: Causal Trees

Athey and Imbens, arxiv:1504.01132v3

- A CART-like or random forest-like algorithm but with an altered split criterion for estimating heterogeneous treatment effects

- Choose the split variables and split points from the observable covariates to maximize

\[
\frac{1}{N} \sum_i \hat{\tau}^2
\]

where

\[
\hat{\tau} \equiv \hat{\mu}(T, x) - \hat{\mu}(C, x)
\]
A More Advanced Model: Causal Trees

Athey and Imbens, arxiv:1504.01132v3

- A CART-like or random forest-like algorithm but with an altered split criterion for estimating heterogeneous treatment effects

- Choose the split variables and split points from the observable covariates to maximize

\[
\frac{1}{N} \sum_i \hat{\tau}^2
\]

where

\[
\hat{\tau} \equiv \hat{\mu}(T, x) - \hat{\mu}(C, x)
\]
Persuasion Model Evaluation
Model Evaluation and Selection

- How do we know if our model is doing a good job?
- Can we design robust checks like we have for regular supervised learning to tell us if our model is working well and to help us choose between different types of models?
Model Evaluation and Selection

- Using a holdout set or cross-validation, get a set of out-of-sample treatment effect scores from a given model.
- Quantile those scores and calculate the true ATE within each quantile.
- Check that those predictions order well and see how they compare to the average predictions in each quantile.
Model Evaluation and Selection

- The uplift curve represents the incremental gain from using the model to target effort or outreach.

- Similar to the quantile plot, rank observations by predicted ATE and compare to actual ATE in each group, blue line.

- Compare this to randomly ordering observations, yellow line.
Model Evaluation and Selection

- The uplift curve represents the incremental gain from using the model to target effort or outreach.
- Similar to the quantile plot, rank observations by predicted ATE and compare to actual ATE in each group, blue line.
  - Compare this to randomly ordering observations, yellow line.

6.5% gain from targeting top 50% of scores vs 2.5% from randomly targeting same number of people
Model Evaluation and Selection

- The **qini coefficient** is analogous to the area under the ROC curve (AUC) for supervised learning.
- A single metric we can use to compare models fit to the same task.
- Scale matters, so we can’t use to compare models in absolute terms.
Model Evaluation and Selection
How We Use Persuasion Modeling at Civis
We built a scientific understanding of each voter.

Our data science targeted voters through paid media, direct mail, social media, communications and fundraising.

Our data science directed decision makers’ strategies and tactics.

We ran the first individualized presidential campaign.
Traditional Social Science Research

Hacking Skills

Math & Statistics Knowledge

Data Science

Substantive Expertise

Social Science Methodology

Danger Zone!

Econometrics

Traditional Research

Traditional Social Science Research
Case Study I: TV Promotional Ad Effectiveness for the Bravo Network
Bravo and Civis partnered to identify swing viewers and to understand how to best persuade them.

**Key Business Questions**

1. Who are Bravo’s “Swing Viewers”?
2. Where or how can we reach them without alienating core viewers?
3. What messaging tone convinces them to spend more time with Bravo?
4. Do different sets of “Swing Viewers” react differently to Bravo’s creative approaches?
We tested five Après Ski promos with different messaging hooks to measure how each piece of creative could increase tune-in.

- **Humor**: Lighthearted moments of the cast in different provocative or comical situations.
- **Luxury**: Lifestyle moments of the wealthy guests interacting with each other + the cast.
- **Altitude**: The “work hard/play hard” professional and personal dichotomy of the lodge staff.
- **Character**: Profile of each of the cast members that displays their personalities and interactions with one another.
- **Attitude**: Displaying moments of conflict and drama between cast members.
We created two meaningful metrics about support for the brand and likelihood to be persuaded by the promo.
We combined the persuasion scores and our Bravo affinity scores to understand how to isolate “swing viewers”
We combined the persuasion scores and our Bravo affinity scores to understand how to isolate “swing viewers”

These People Will Likely Tune In Anyways Because of their High Support
We combined the persuasion scores and our Bravo affinity scores to understand how to isolate “swing viewers”.

These People Won’t Watch No Matter What

Persuasion Effect

Level of Bravo Affinity
We combined the persuasion scores and our Bravo affinity scores to understand how to isolate “swing viewers”.

Bravo’s Swing Viewers: A Casual but Persuadable Group of 22 Million Adults
Case Study II: Persuasion in the 2016 Election Cycle
In early 2016, we conducted a randomized controlled message test for a client using tens of thousands of responses in 14 states around the country.

- We tested 3 messages: “women’s health”, “the future of Medicaid and Social Security”, and “tax cuts for the wealthy”.

- We averaged the persuasion scores from “the future of Medicaid/Social Security” and “tax cuts for the wealthy” messages for a general “economy persuasion score”. We averaged all three scores to create a “generic persuasion score”.

In August (8 months later), we conducted a follow-up randomized, controlled video ad test in Pennsylvania, which allowed us to validate these persuadable segments.
### Remarkable result

Our persuasion scores reliably predicted the movement of opinion 8 months later in a completely different context.

The top quartile of people that the model predicted to be most persuadable moved 3x-4x as much as the least persuadable people.
Case Study III: TV Promotional Ad Effectiveness from Observational Data
TV Ad Effectiveness from Observational Data

- With purely observational data on who has seen an advertisement, we don’t have nice randomization like we do in a randomized controlled trial.
  - Maybe people who saw the advertisement are systematically different than those who didn’t.

- It’s possible to use techniques like propensity score matching from the causal inference literature to correct for this.
  - We construct a matched “synthetic” control group who looks like the treatment group in their viewership behavior but just happened to miss the advertisement that we’re studying.
Observational Study

- Saw Ad
- Didn’t See Ad
- Propensity Model
- Treatment
- Control
- Measure Viewership
Observational Study

Saw Ad → Propensity Model → Measure Viewership

Didn’t See Ad → Control

Model predicting exposure to the ad
Propensity Model

Saw Ad

Propensity Model

Didn’t See Ad

Control

Discard the observations that are too “different” from the ad viewers

Measure Viewership
A True Crime Series

56, female, 63K/yr

![Graph showing treatment effect and lift comparison.](image-url)
A Family Reality Series

- **Graph 1:** Bar chart showing the predicted treatment effect decile with actual and predicted values.
  - Y-axis: Treatment Effect (0% - 4%)
  - X-axis: Predicted Treatment Effect Decile

- **Graph 2:** Lift chart comparing model and random population fraction.
  - Y-axis: Lift
  - X-axis: Population Fraction (%)
A Family Reality Series

52, Black, from the southwest, 72K/yr, not a cat person
These commercials don’t seem to convince most young people...
Parting Thoughts
Use **persuasion modeling** when you need to optimally allocate treatments or interventions to achieve some outcome.
We’ve open sourced some of our data science tools and plan to release a few of the things we discussed today. Watch GitHub or our blog.

GitHub: github.com/civisanalytics
Website: civisanalytics.com/open-source/
Thanks!

@MichelangeloDA            @wlattner