Uplift Modeling in Algorithmic Marketing

William Wong

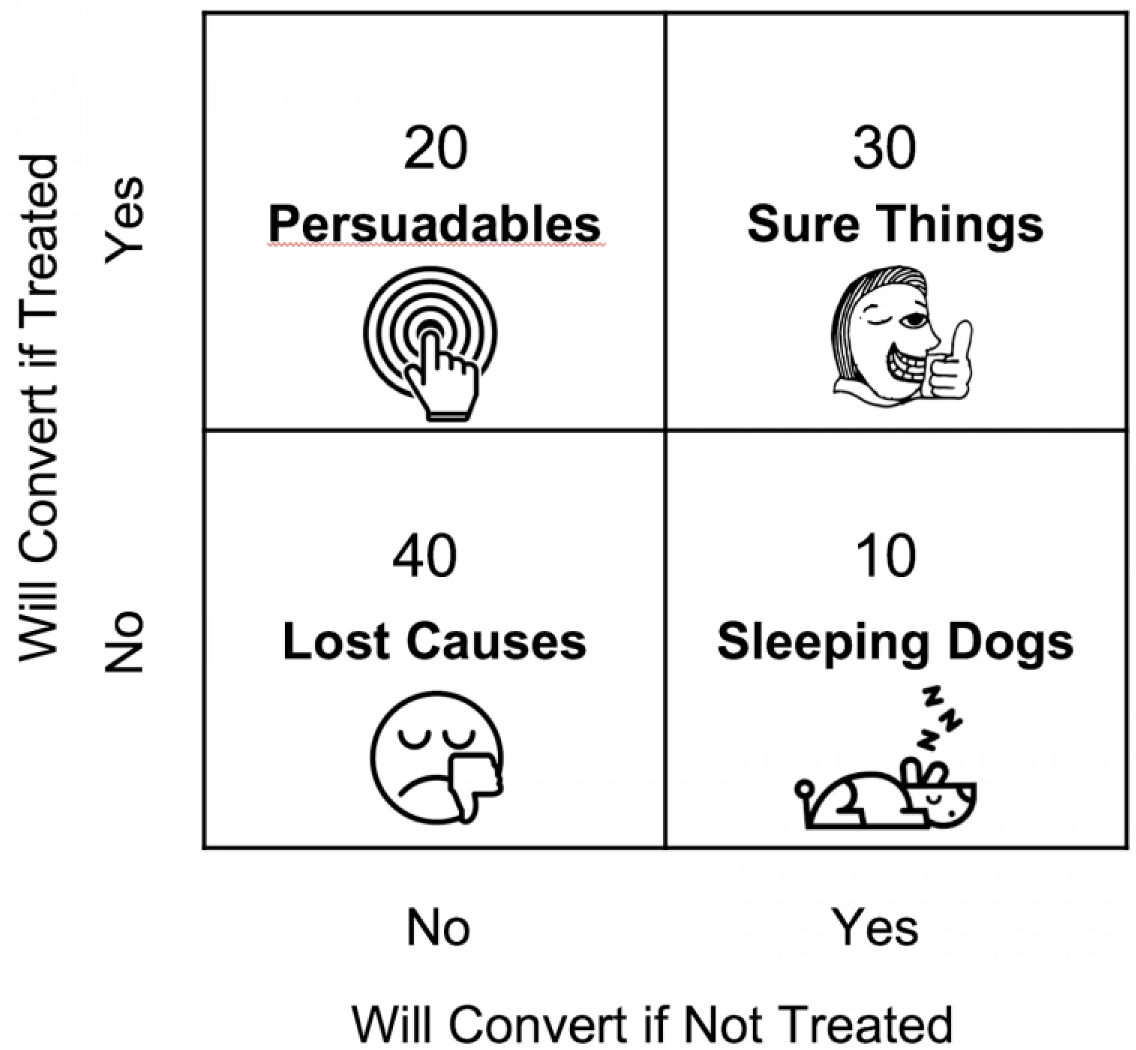
January 4, 2022

Outline

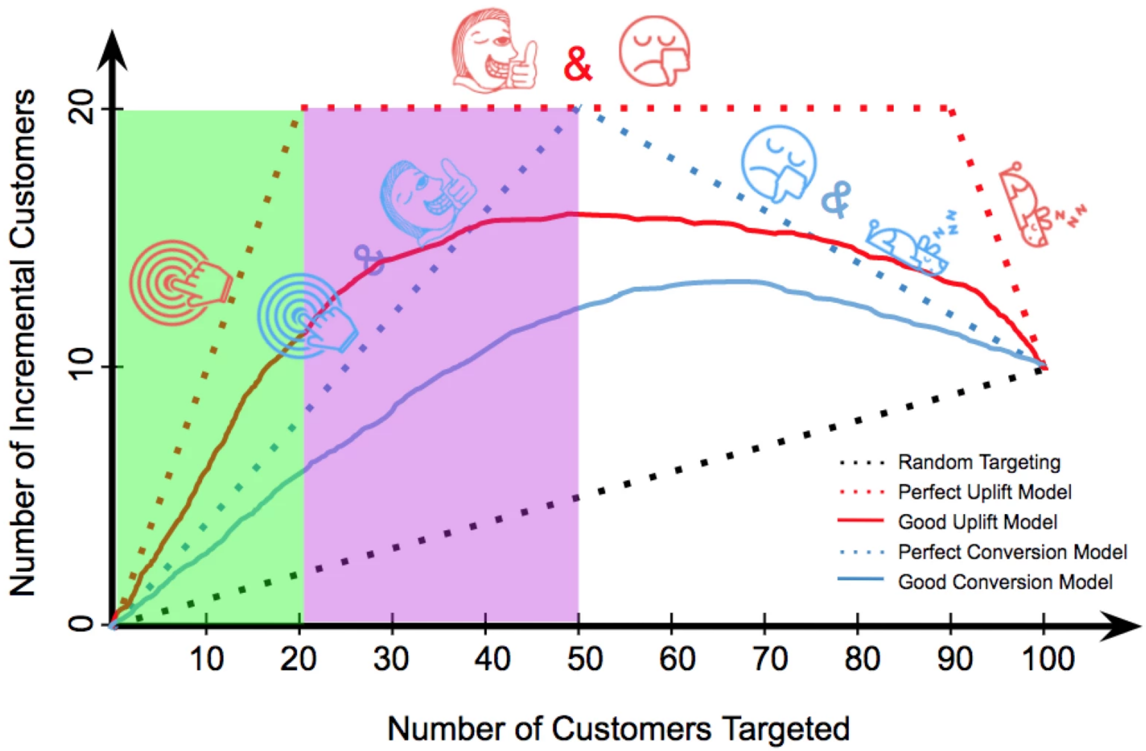
- Introduction
- Review of causal inference
- Three methods of modeling uplift
- Numerical example using CausalML

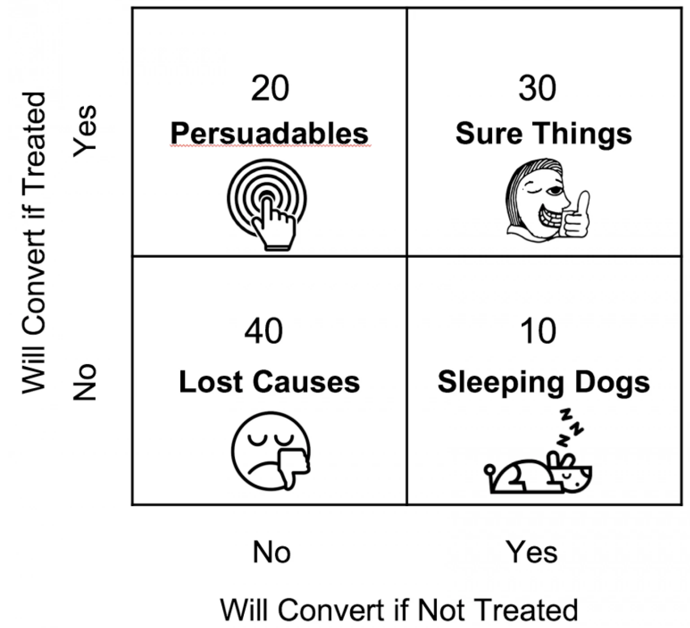
What is Uplift Modeling?

- Uplift modeling refers to the set of techniques used to model the incremental impact of an action or treatment on a customer outcome.
- It is both a causal inference problem and a machine learning one.
- There are 100 customers belonging to 4 segments as shown in the figure below.



Cumulative Uplift





- In the example above, there are 100 customers who belong to the 4 separate segments.
- The use cases for uplift modeling:
  - Target the Persuadables for promotions.
  - Stop reaching out to those who react to the treatment negatively (e.g., the Sleeping Dogs).

Review of Causal Inference

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
T \arrow[r] & Y
\end{tikzcd}
$$

- X is a confounder for the treatment T and the outcome Y.
- $\mathbb{E}[Y^1]$ is the average value of Y if **everyone was treated** with $T = 1$.
- The average treatment effect ATE $= \mathbb{E}[Y^1 - Y^0]$.
- $\mathbb{E}[Y^1 - Y^0] \neq \mathbb{E}[Y|T = 1] - \mathbb{E}[Y|T = 0]$
  - $\mathbb{E}[Y^1 - Y^0]$ is the **average treatment effect**, because it is comparing what would happen if the same people were treated with $T = 1$ versus with $T = 0$.
  - $\mathbb{E}[Y|T = 1] - \mathbb{E}[Y|T = 0]$ is the **observed treatment effect**. Note that it is comparing two different populations of people.
- An example
  - T is COVID vaccination.
  - Y is mortality.
  - X is age.
- Consistency assumption — the potential outcome under treatment $Y^{T=t}$ is equal to the observed outcome if the actual treatment received is $T = t$.
- Ignorability assumption — $\{Y^1, Y^0\} \perp T|X$. Among subjects with the same values of X, we can think of treatment T as being randomly assigned.

Review of Causal Inference (con't)

In a randomized trial, the distribution of X will be the same in both groups since the assignment is random!

$$\begin{tikzcd}
X \arrow[dr] & \\
T \arrow[r] & Y
\end{tikzcd}$$

For observational data,

$$\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
T \arrow[r] & Y
\end{tikzcd}$$

we can match individuals in the $T = 1$ group to individuals in the $T = 0$ group on the covariates X.

Problem Formulation

$$
\begin{tikzcd}
X \arrow[d] \arrow[dr] & \\
T \arrow[r] & Y
\end{tikzcd}
$$

- Paper by Gutierrez and Gérardy, 2016 [link].
- The Conditional Average Treatment Effect for a subgroup in the population

$$\boxed{\text{CATE} = \tau(X) = \mathbb{E}[Y^1|X] - \mathbb{E}[Y^0|X]}\ , \tag{1}$$

where X is a vector of features.

  - $\mathbb{E}[Y|T = t, X]$ references observed data only.
  - $\mathbb{E}[Y|T = t, X] = \mathbb{E}[Y^{T=t}|T = t, X]$ by consistency.
  - $\mathbb{E}[Y|T = t, X] = \mathbb{E}[Y^{T=t}|T = t, X] = \mathbb{E}[Y^{T=t}|X]$ by ignorability.

Comparison of Various Models

Model	**Prediction**
Propensity model	$\Pr(\text{buy}\|T = 0, X)$
Churn model	$\Pr(\text{churn}\|T = 0, X)$
Response model	$\Pr(\text{buy}\|T = 1, X)$
Uplift model	$\Pr(\text{buy}\|T = 1, X) - \Pr(\text{buy}\|T = 0, X)$
Style-affinity model	$\Pr(\text{style} = s\|\text{buy}, X)$
Price-affinity model	$\Pr(\text{price} = p\|\text{buy}, X)$

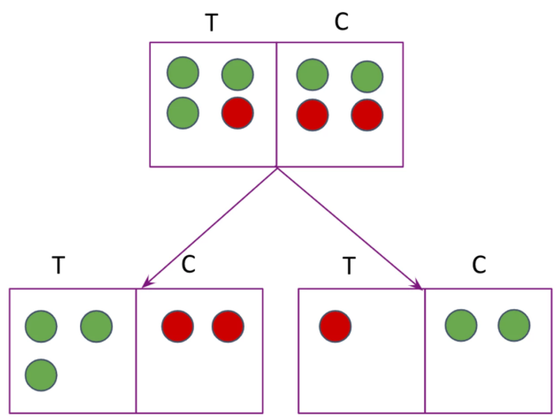
Method 1 — Build Two Separate ML Models

- Estimate $E[Y^1|X]$ and $E[Y^0|X]$ using the treatment group data and the control group data separately.

Method 2 — Class Transformation

- Define a new outcome variable $Y^* = Y^1 \frac{T}{\Pr(T=1|X)} - Y^0 \frac{1-T}{1-\Pr(T=1|X)}$.
- Can show that $\mathbb{E}[Y^*|X] = \text{CATE} = \tau(X)$.
- Build a model to estimate $\mathbb{E}[Y^*|X]$.

Method 3 — Direct Modeling using a Decision Tree



- **Modify**(!) existing ML algorithms to model the uplift.
- There are 8 data points in a given tree node, with 4 instances in the treatment group and 4 instances in the holdout. Three out of the 4 customers in the treatment group converted (green circles), and 2 out of the 4 customers in the holdout group converted (red circles).
- For the best split at a given node in the tree, we want to maximize the gain of the divergence between the outcome class distributions between treatment and control [link].
- The left child node contains the Persuadables only. Everyone in the treatment group converted, and no one in the control group converted.
- The right child node is just the opposite; it contains the Sleeping Dogs who generate negative value when they receive treatment.

Uplift Tree and Random Forests using CausalML

- CausalML [link] is an open-source Python package from Uber.
- Jupyter notebook [link].
- The synthetic dataset contains columns
  - `treatment_group_key`: Each row belongs to one of the **four** groups — `control`, `treatment1`, `treatment2`, and `treatment3`. There are 1,000 rows for each group.
  - 19 features.
  - `conversion`: 0 or 1.
- Uses `UpliftRandomForestClassifier()` as the model.
- CausalML has the `plot_gain()` function which calculates the uplift curve given a DataFrame containing the treatment assignment, observed outcome, and the predicted treatment effect.