Causal KNN

Seminar Applied Predictive Analytics

Maximilian Kricke
Tim Peschenz

Chair of Information Systems
Humboldt Universität zu Berlin

27.06.2019

Outline

Course Framework

Methodology Perspective

Application Perspective

- Paper: "Heterogeneous Treatment Effects and Optimal Targeting Policy Evaluation" (Hitsch and Misra 2018)
- Potential Outcomes Framework and the Fundamental Problem of Causal Inference: *We never observe the individual treatment effect but only the outcome corresponding to the assigned treatment.*

Course Framework of our Topic

- Method to somehow overcome this fundamental problem
- General parameter tuning technique for uplift models
- Applicable tool to evaluate different targeting policies using only one data set

Business Framework

- **Imagine you run a marketing campaign**
- How do you measure if a campaign was successful or not?
- To optimize a marketing campaign, you should decide based on customer level:
  - A customer should be targeted if the individual profit contribution is greater than the targeting cost for this individual.

It is shown that the CATE on an individual basis is sufficient to construct an optimal targeting policy.

But how can we measure the (unobservable) individual CATE?

Measuring the individual CATE

We need to make three assumptions to the data:

- Unconfoundedness: $Y_i(0), Y_i(1) \perp W_i | X_i$ (random assignment of the treatment within each subgroup of the population with identical features $X_i = x$)
- No overlap: $0 < e(x) < 1$
- Stable Unit Treatment Value Assumption (SUTVA): No social interactions or equilibrium effects

If these assumptions are satisfied, we can compute the CATE via:

$$\tau(x) = \mathbb{E}[Y_i | X_i, W_i = 1] - \mathbb{E}[Y_i | X_i, W_i = 0]$$

Hence we are calculating the mean difference between the outcomes of treated and untreated units with identical features.

But do we even have individuals with the same features?

Outline

The Causal KNN Approach

- direct estimation of the conditional average treatment effect $\hat{\tau}_i$ for all observations i, by comparing the mean outcome of the nearest neighbours
- selecting nearest neighbours based on the covariates X_i using the euclidian distance measure
- compare outcomes Y of k treated $Y(w=1)$ and k untreated $Y(w=0)$ nearest neighbours of unit i, to derive an estimation $\hat{\tau}_i$ of the conditional average treatment effect τ_i

$$\hat{\tau}_K(x) = \frac{1}{K} \sum_{i \in \mathcal{N}_K(x,1)} Y_i - \frac{1}{K} \sum_{i \in \mathcal{N}_K(x,0)} Y_i$$

The Causal KNN Approach

Algorithm 1 Causal KNN Algorithm

procedure $\textsc{CausalKNN}(k, y, w, x)$

Input:

$k \rightarrow$ number of nearest neighbours

$y \rightarrow$ vector with outcome values

$w \rightarrow$ vector with treatment status

$x \rightarrow$ matrix with covariates

Begin

1. select nearest neighbours for each observation i based on the covariates x
2. separate k treated $(w_i = 1)$ and k untreated $(w_i = 0)$ nearest neighbours

for each observation i **do**

calculate mean of $y_k(w = 1)$ and $y_k(w = 0)$

$uplift_i$ = mean of $y_k(w = 1)$ - mean of $y_k(w = 0)$

End

Causal KNN Example

- k value has to be set in advance
- example table with necessary input information

i	**W**	**Y**	X_1	X_2	$\cdots$	X_m
1	1	10	x_{11}	x_{12}	$\cdots$	x_{1m}
2	1	20	x_{21}	x_{22}	$\cdots$	x_{2m}
3	0	25	x_{31}	x_{32}	$\cdots$	x_{3m}
4	0	10	x_{41}	x_{42}	$\cdots$	x_{4m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	1	30	x_{n1}	x_{n2}	$\cdots$	x_{nm}

$$\hat{\tau}_K(x) = \frac{1}{K} \sum_{i \in \mathcal{N}_K(x,1)} Y_i - \frac{1}{K} \sum_{i \in \mathcal{N}_K(x,0)} Y_i$$

But how should we decide on the optimal value of k?

Outline

Transformed Outcome

- **Idea**:
  Define a proxy for the true CATE $\tau(X_i)$ to apply the MSE-Criterion:
  $\mathbb{E}[(\tau(X_i) - \hat{\tau}_K(X_i))^2]$
- Define the transformed outcome as:

$$Y_i^* = W_i \cdot \frac{Y_i(1)}{e(X_i)} - (1 - W_i) \cdot \frac{Y_i(0)}{1 - e(X_i)} \tag{1}$$

$$MSE_b = \frac{1}{|\tau_b|} \sum_{i \in \tau_b} (Y_i^* - \widehat{\tau_b}(X_i))^2 \tag{2}$$

- We can rewrite this to:

$$Y_i^* = \frac{W_i - e(X_i)}{e(X_i)(1 - e(X_i))} Y_i^{obs}$$

to observe it for all units i.

Justification for the Transformed Outcome:

$$
\begin{aligned}
Y_i^* &= W_i \cdot \frac{Y_i(1)}{e(X_i)} - (1 - W_i) \cdot \frac{Y_i(0)}{1 - e(X_i)} \quad \Big| \quad \mathbb{E}[\cdot | X_i = x] \\
&= \mathbb{E}[W_i | X_i = x] \cdot \frac{\mathbb{E}[Y_i(1) | X_i = x]}{e(X_i)} - \mathbb{E}[1 - W_i | X_i = x] \cdot \frac{\mathbb{E}[Y_i(0) | X_i = x]}{1 - e(X_i)} \\
&= e(X_i) \cdot \frac{\mathbb{E}[Y_i(1) | X_i = x]}{e(X_i)} - (1 - e(X_i)) \cdot \frac{\mathbb{E}[Y_i(0) | X_i = x]}{1 - e(X_i)} \\
&= \mathbb{E}[Y_i(1) | X_i = x] - \mathbb{E}[Y_i(0) | X_i = x] \quad \Big| \quad \textbf{Unconfoundedness} \\
&= \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x] \\
&= \tau(X)
\end{aligned}
$$

Hence, the transformed outcome is an unbiased estimate of the CATE and can therefore be used as the desired proxy.

Transformed Outcome Loss

- Now we can define the transformed outcome loss as:

$$\mathbb{E}[(Y_i^* - \hat{\tau}_K(X_i))^2]$$

- The value for k that minimizes the outcome loss is optimal for our estimation

$$\arg\min_K = \mathbb{E}[(Y_i^* - \hat{\tau}_K(X_i))^2] \tag{3}$$

Outline

Application: Data Presentation

	recency	history_segment	history	mens	womens	zip_code	newbie	channel	segment	visit	conversion	spend
1	10.00	2) $100 - $200	142.44	1	0	Suburban	0	Phone	Womens E-Mail	0	0	0.00
2	6.00	3) $200 - $350	329.08	1	1	Rural	1	Web	No E-Mail	0	0	0.00
3	7.00	2) $100 - $200	180.65	0	1	Suburban	1	Web	Womens E-Mail	0	0	0.00
4	9.00	5) $500 - $750	675.83	1	0	Rural	1	Web	Mens E-Mail	0	0	0.00
5	2.00	1) $0 - $100	45.34	1	0	Urban	0	Web	Womens E-Mail	0	0	0.00
6	6.00	2) $100 - $200	134.83	0	1	Suburban	0	Phone	Womens E-Mail	1	0	0.00

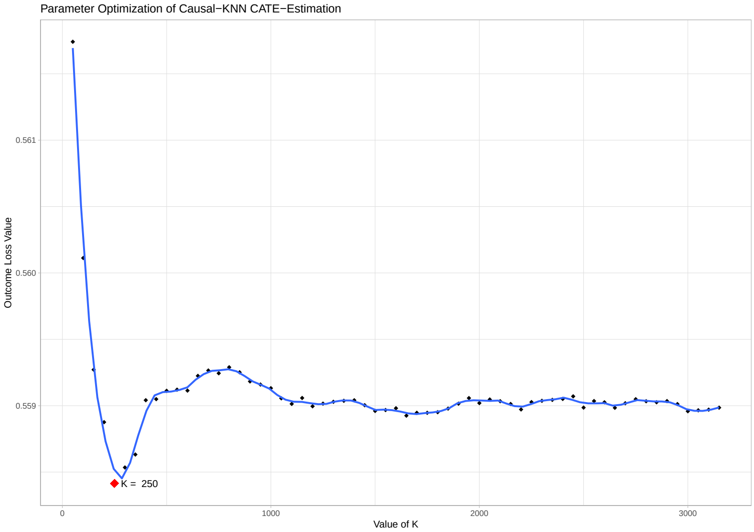
- Data base that captures 64.000 customer transactions
- Treatment Variable "segment"
- Outcome variables: "visit", "conversion" (binary) and "spend" (numeric)

Application: Process

- Calculate individual CATE for all observations in the data using Causal KNN
- Select optimal K value by finding the minimal transformed outcome loss value
- choosing parameter K to minimize expected transformed outcome loss similar to minimize expected mean squared error (MSE)

- parameter tuning for different Uplift Modeling Technique
- Causal Tree Model Tuning

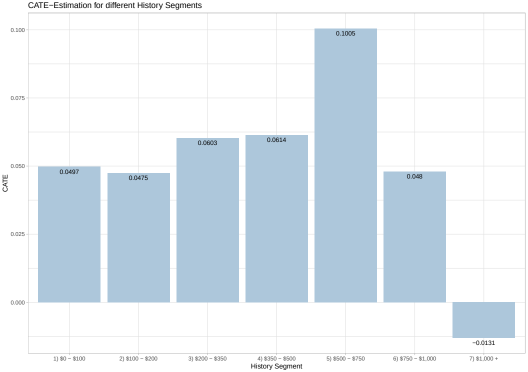
Application: Results

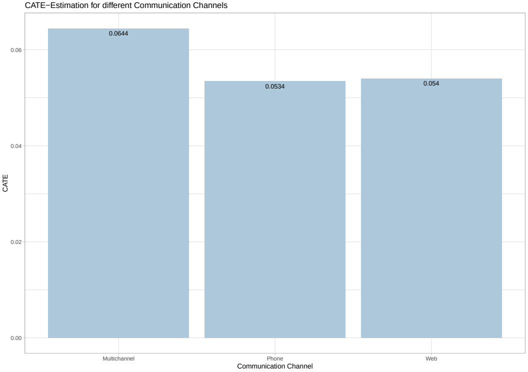


Application: Results

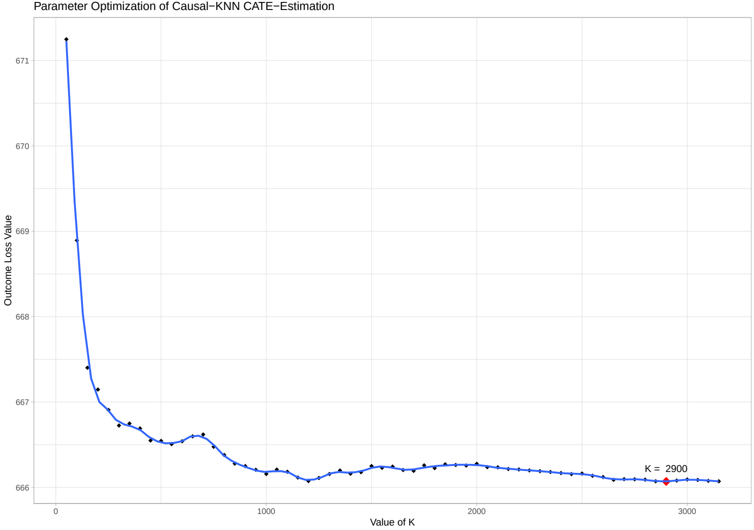
Outcome Variable: Visit (binary)

- ATE $= 0.05497$
- ATT $= 0.05513$





Application: Results



Application: Results

Outcome Variable: Spend (numeric) in $ (US)

- ATE $= 0.31321$
- ATT $= 0.31205$