MSOM 2026, Harvard

Decision-Aware Segmentation

Ke Li
PhD Candidate in Decision Sciences
INSEAD

Ke Li

Sandeep Chandukala

Ernst Osinga

Spyros Zoumpoulis

Customer Segmentation Drives Business Decisions

- Department stores (e.g., Sephora): coupon campaigns
- E-commerce, online retailers (e.g., Amazon): ad campaigns
- Banks (e.g., American Express): discount offers

SEPHORA

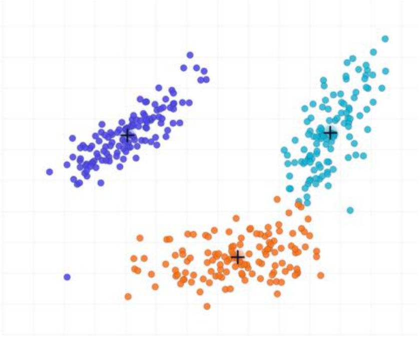
amazon

AMERICAN EXPRESS

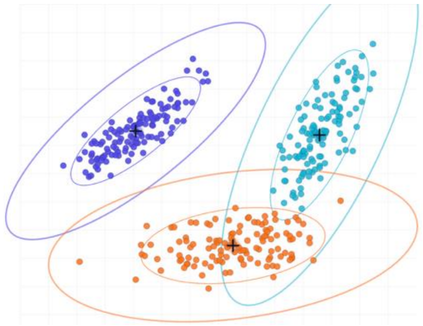
But: Are We Segmenting for the Right Objective?

Widely-used segmentation algorithms: optimizing statistical fit/separation

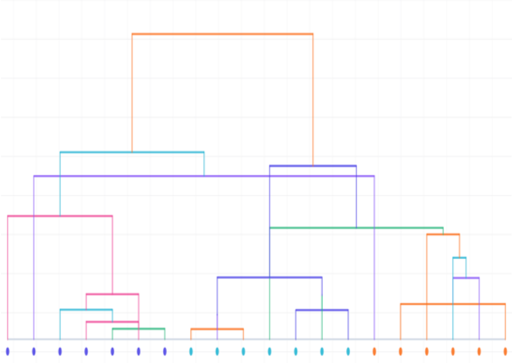
Centroid-based clustering
(e.g., K-means)



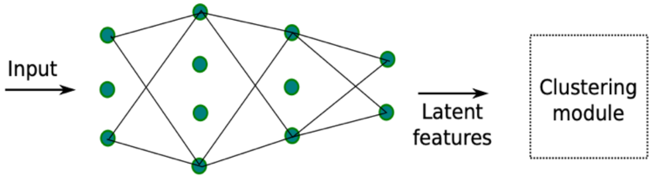
Model-based clustering
(e.g., GMM)



Hierarchical clustering



ML/DL-based clustering



Segmentation Should Optimize Decision Outcomes

Descriptive Segmentation → **Targeting Policies**

Optimizing for statistical fit

Optimizing for decisions/profits

- **Uplift modeling** (Ascarza, 2018; Lemmens and Gupta, 2020; …)
- **Policy learning** (Manski, 2004; Tetenov, 2012; Kitagawa, 2018; Athey and Wager, 2021; …)
- **Decision-aware learning** (Mandi et al., 2024; Elmachtoub et al., 2020, 2022; Vanderschueren et al., 2022; …)

Prescriptive Segmentation

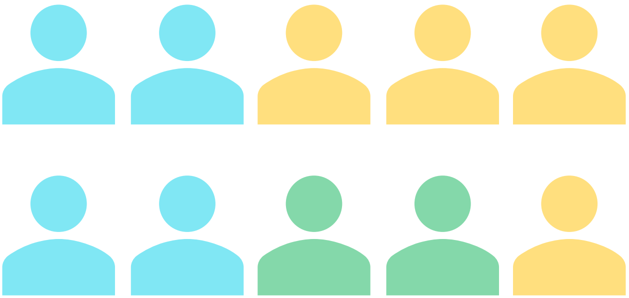
<u>**Our work**</u>: optimizing segmentation for targeting decisions/profits

Incorporate managerial objectives

- **Prescriptive segmentation** (Liu et al., 2010; Aouad et al., 2023; Zhang and Misra, 2024; Cui and Hamilton, 2026)

Research Questions

How should firms segment to optimize decision outcomes?

How many segments	**How to segment**
How should firms choose the right number of segments? 2 : 3 : 	Given a number of segments, how should customers be partitioned?

Running Example

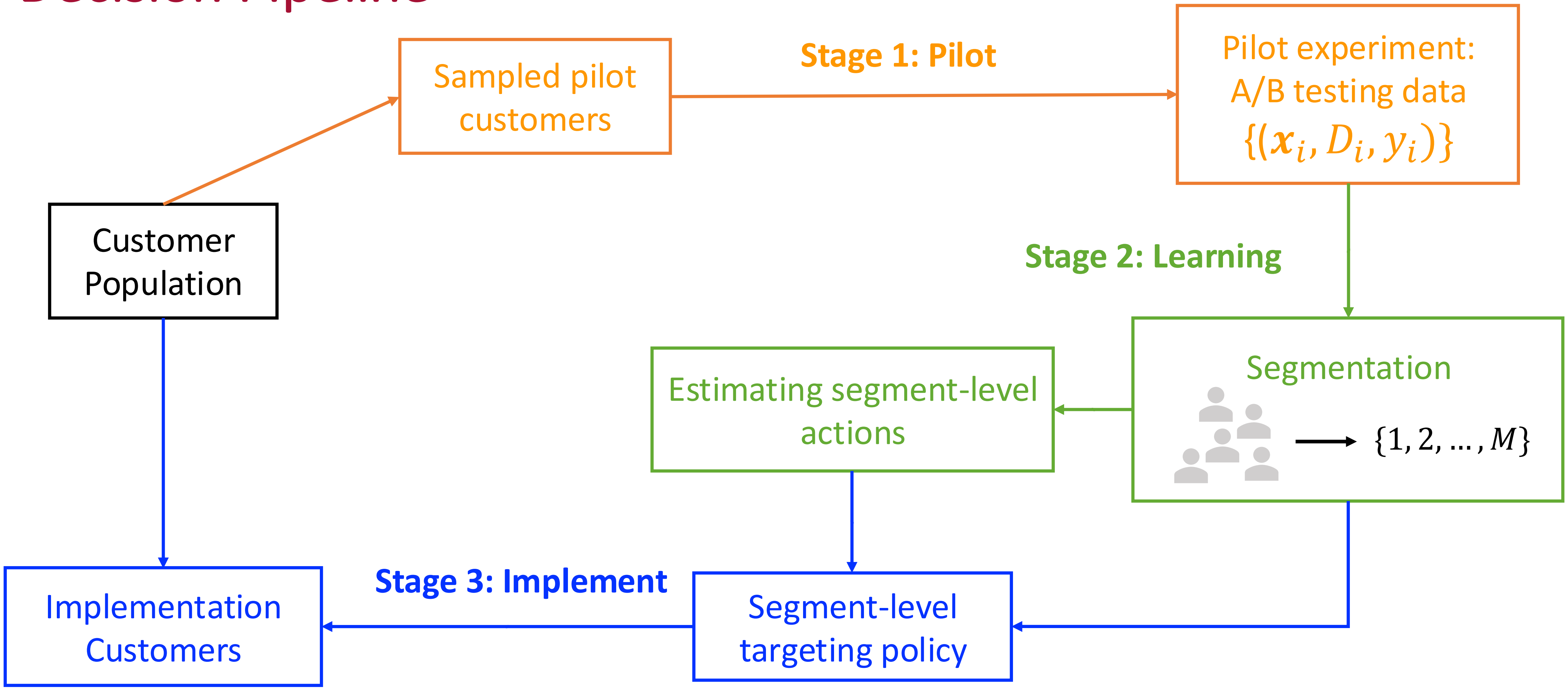
- Suppose an e-commerce retailer wants to determine how to allocate Ad A and Ad B across customers to maximize conversions.

- The firm can choose to show Ad A (D=0) or Ad B (D=1) to each customer.

- Fully individualized? Practical limitations
  - over-fitting due to insufficient data; costly to obtain data
  - difficult to operationalize

- The firm wants to deploy a **segment-level targeting** instead.

Decision Pipeline

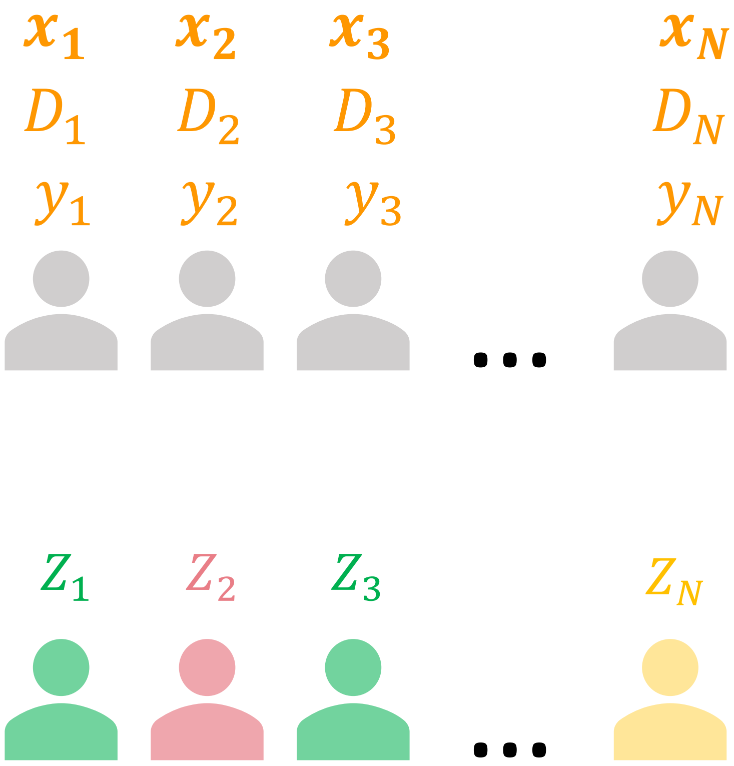


How to segment customers to maximize implementation outcomes?

Problem Setup

(Observables: pilot data)

- Each customer i with a feature vector $\boldsymbol{x}_i \in \mathbb{R}^d$
- treatment $D_i \in \{0,1\}$, and realized outcome y_i



(Unobservables: latent structure)

- K latent segments: each segment k characterized by (f_k, θ_k) .

(Data generation process)

- The realized response y_i for customer i follows a logistic model:

$$y_i = \boxed{\sigma\left(\underbrace{f_{Z_i}(\boldsymbol{x}_i)}_{\text{baseline}} + \boxed{\theta_{Z_i}} \cdot D_i \right)} \qquad \sigma(x) = \frac{1}{1+e^{-x}}$$

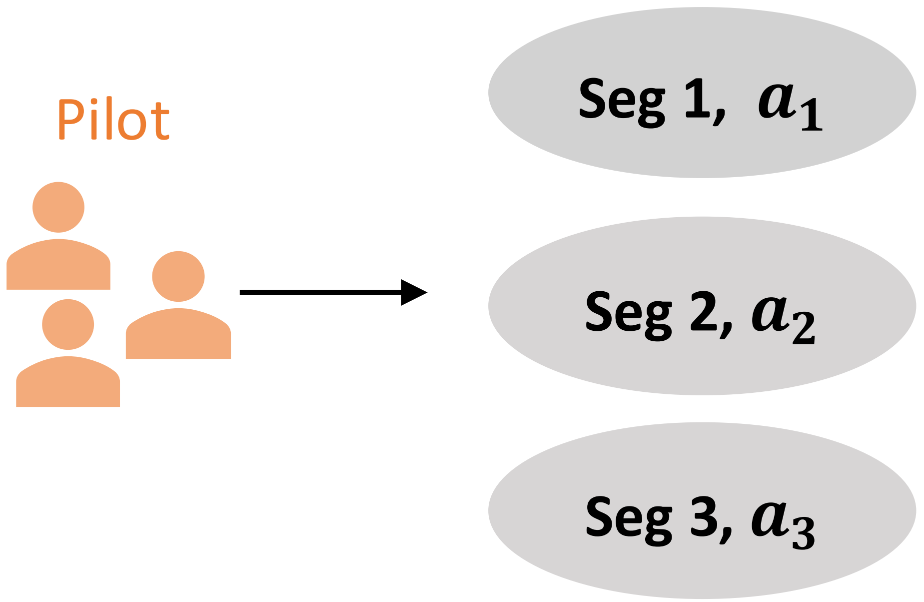
oberved by managers

treatment parameter

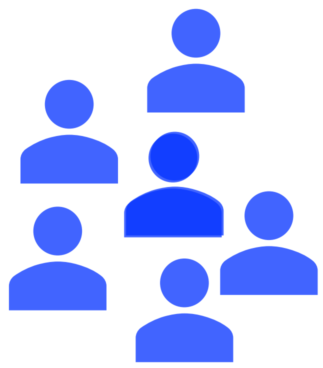
Not known by managers

Manager’s Optimization Problem

The manager learns a segmentation ϕ from the pilot data $\{(\boldsymbol{x}_i, D_i, y_i)\}$, and estimates targeting actions for each segment.



$$\boxed{r_j = \sigma\left(f_{Z_j}(\boldsymbol{x}_i) + \theta_{Z_j} \cdot a_j \right)}$$



The manager seeks a segmentation ϕ that maximizes the sum of probabilities

$$\max_{\phi} \sum_{j \in implement} r_j$$

Optimization Proxy

Not observable during the learning stage!

The manager's optimization goal:

$$\max_{segmentation\ \phi} \sum_{j \in Implement} \boxed{r_j}$$

Instead, we optimize over an estimate $\hat{r}$ of the outcome

$$\max_{segmentation\ \phi} \sum_{j \in Implement} \hat{r}_j$$

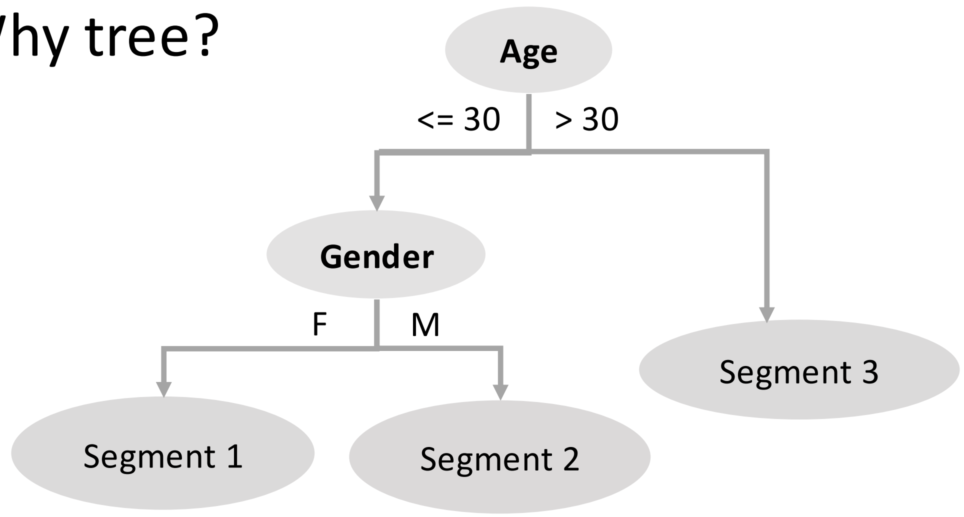
We use doubly-robust (DR) score, a per-customer, per-action estimate of counterfactual outcomes (Dudík, M. et al., 2011).

$$(\text{customer } j, \text{ action } a_j) \longrightarrow \boxed{\text{DR estimator}} \longrightarrow \hat{r}_j$$

Decision-Aware Segmentation (DAS)

A framework that produces tree-based segmentation by directly maximizing the downstream outcomes.

Why tree?



- Natural segmentation structure (leaf = segment)
- Interpretable
- Non-linear relationships
- Easy to allocate

Decision-Aware Segmentation (DAS)

How many segments	How to segment
	Algorithm 1 DAST (Decision-Aware Segmentation Tree) A tree-based algorithm to segment customers for a given number of segments

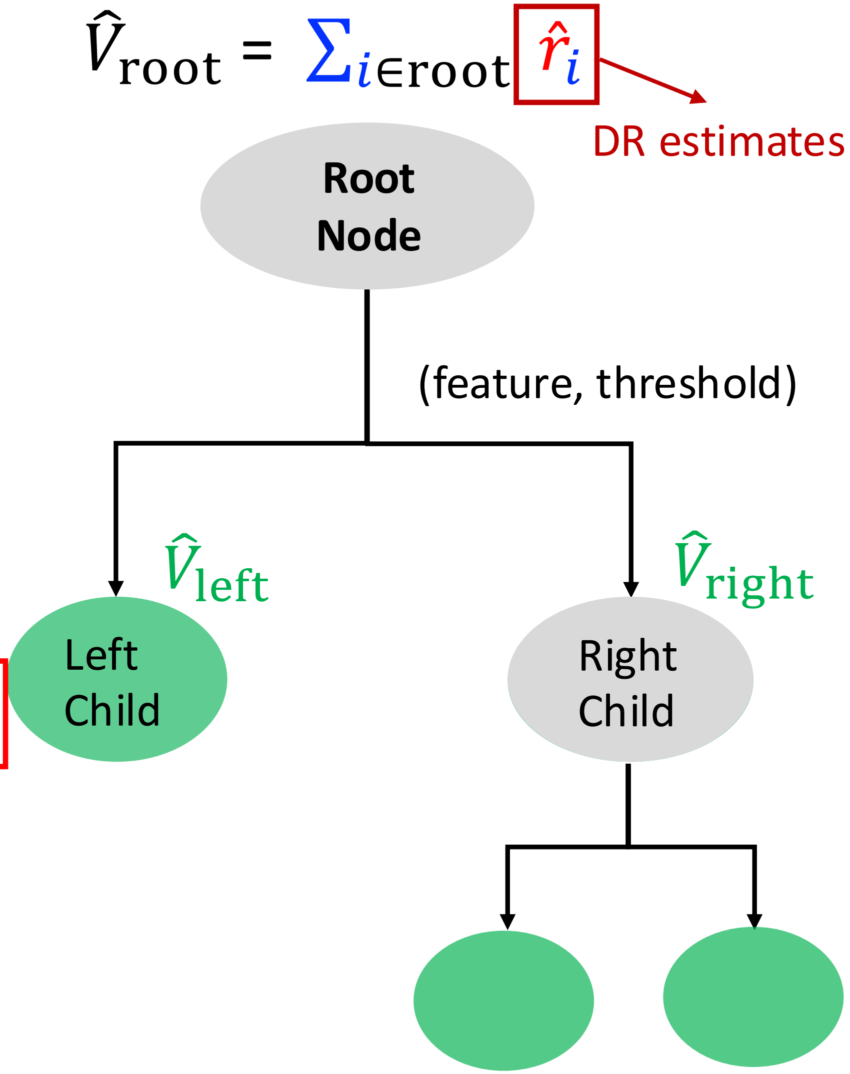
Algorithm 1 – Segmentation DAST

Build on MST (Aouad et al., 2023) and Policytree (Sverdrup et al., 2020),
- better aligned with managerial objectives
- more computationally efficient

Algorithm procedure:

1. Grow a segmentation tree:
   - For each current leaf, evaluate all feature-threshold pairs by policy-value gain $\boxed{\Delta V = \hat{V}_{\text{left}} + \hat{V}_{\text{right}} - \hat{V}_{\text{parent}}}$
   - Split the leaf node that yields the largest gain $\Delta V_{\max}$
   - Repeat the above steps until the tree grows to the desired number of leaves M

2. Estimate leaf-level actions



Decision-Aware Segmentation (DAS)

How many segments	How to segment
Algorithm 2 DAMS (Decision-Aware Model Selection) A model selection method for choosing the number of segments	**Algorithm 1 DAST (Decision-Aware Segmentation Tree)** A tree-based algorithm to segment customers for a given number of segments

Algorithm 2 – Model Selection DAMS

Algorithm procedure:

- Specify candidate segment counts, e.g., $\mathcal{M} = \{1, 2, 3, .., 10\}$
- For each $M \in \mathcal{M}$, learn a segmentation ϕ_M (Algorithm 1 - DAST) and estimate targeting actions for customers in a held-out validation set
- Evaluate each segmentation ϕ_M,

$$\hat{V}(M) = \sum_{j \in validation} \hat{r}_j$$

- Select the M^* with the highest estimated value

$$M^* \in \underset{M \in \mathcal{M}}{\operatorname{argmax}} \hat{V}(M)$$

We implement a k-fold cross-validation version

Theoretical Guarantees

Question 1: Can DAST scale to millions of customers?

In math language: Time complexity of DAST: $O(d \cdot N \log N)$

In human-friendly language: DAST scales efficiently (near-linearithmic) to large datasets

Theoretical Guarantees

Question 2: Will DAST choose the right split, or just chase random noise?

In math language:

$$V(h^*) - V(\hat{h}) \leq C \cdot \sqrt{\log(\frac{16 \cdot |\mathcal{H}|}{\delta})}$$

In human-friendly language:

With high probability, the split chosen by DAST achieves a performance close to the best possible candidate under perfect information.

Theoretical Guarantees

Question 3: How many validation samples do we need to reliably identify the true number of segments?

In math language:

$$N \geq C \cdot \frac{1}{\eta^2} \log(\frac{2 \cdot |\mathcal{M}|}{\delta})$$

In human-friendly language: The required validation sample size grows **logarithmically** with the number of candidate segmentations.

Numerical Studies: Benchmark Methods

- List of Benchmark Methods (~10)

Segmentation Methods	HTE-based Methods
➢ **K-Means** (Hartigan and Wong, 1979) ➢ **GMM** (Dempster et al., 1977) ➢ **Clusterwise-Linear Regression** (DeSarbo and Cron, 1988) ➢ **Market Segmentation Trees** (Aouad et al., 2023)	➢ **Meta-learners**, including **T-learner, S-learner, X-learner, and DR-learner** (Künzel et al., 2019; Kennedy, 2023) ➢ **Causal Forest** (Wager and Athey, 2018) ➢ **Policytree** (Sverdrup et al., 2020)

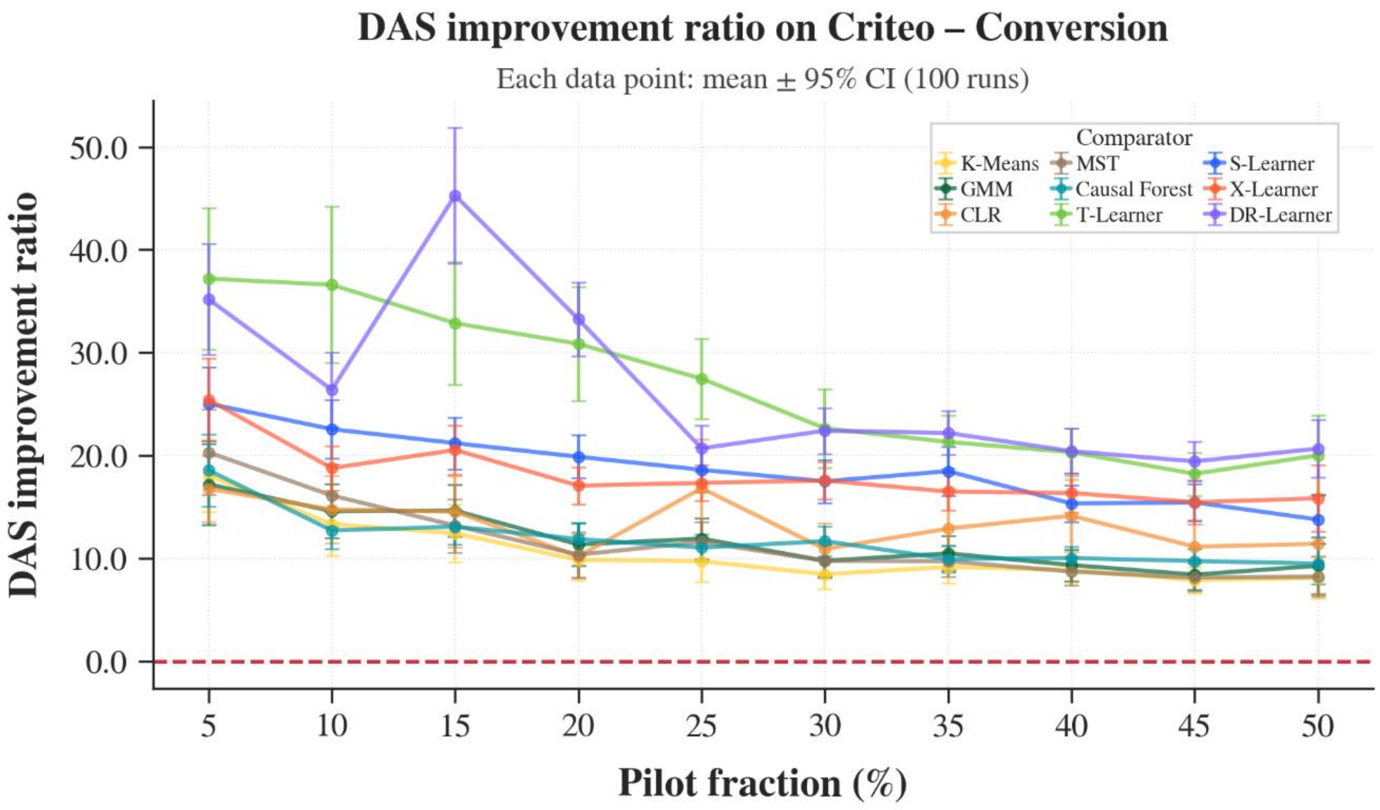
- Compare DAS with comparators on both
  - Synthetic datasets
  - Real-world datasets (e.g., Hillstrom Email, Criteo Uplift)

Real-world Result - Improvement on Targeting Outcomes

CRITEO

Online advertising A/B tests (RCT)

- **N**: 140,000
- **X**: 12 features
- **Treatment**: binary (0/1)
- **Y**: conversion rate (very sparse, ~0.3%)



DAS improves the conversion rate by 10~30% on real-world dataset.

Open Source (Python + R Package coming soon!)

- One-step installation: `pip install XXX` or `install.packages("XXX")`
- Minimal setup: 4 lines of code

```
from das import DASTree

tree = DASTree(M=3)
tree.fit(X, y, D)

actions  = tree.predict_action(X_new)
```

Prescriptive actions

- Support multi-action settings
- Easy integration with existing targeting pipelines

Takeaways: Optimizing segmentation for the decisions

- **A new perspective:** Segmentation is not just a statistical problem, but also a decision problem.
- **A new solution:** DAS directly optimizes segmentation for downstream targeting decisions.
- **A deployable tool:** An open-source Python/R package for practitioners.

Thank You & Questions

INSEAD The Business School for the World®

My research interests:

- Operations, machine learning, decision-aware learning
- Human–AI collaboration, AI for Science

Open to discussions and collaboration! kelichloe.github.io