Targeting and Privacy in Mobile Advertising

Rafieian and Yoganarasimhan

University of Washington

Mobile Adoption and Usage

- Smartphones are increasingly popular
  - 2 Billion users worldwide
  - Avg. user spends 3.3 hours/day
  - Bulk of usage through apps
  - In 2016, Internet usage via smartphones and tablets exceeded desktop for the first time worldwide
- Mobile advertising
  - Worldwide revenue of 183 billion USD in 2018
    - Predicted to exceed 200 billion USD in 2019
  - Largest share of total digital ad spend
    - Over 68% of total digital ad revenues

In-App Advertising

- Common app monetization strategy
  - App developers can earn money through ads
- Excellent user tracking and targeting properties
  - Advertisers have access to a device ID — IDFA in Apple and AAID in Android
  - Persistent unless re-set by user
  - Used to stitch user data across sessions, apps, and ads

Targeting and Privacy Trade-off

- Better tracking techniques improve behavioral targeting
  - Increased efficiency in the market
  - However, this has led to privacy concerns among users
- Part of broader debate over consumer tracking and privacy
  - Advertisers: fewer protections, behavioral tracking tools
  - Consumers: higher privacy, limits on targeting
  - Regulators: e.g., GDPR by EU — balance profitability motives with consumer protection. Self-regulation?

Revenue-Efficiency Trade-off

- Higher efficiency does not lead to higher revenues
  - Fat-tailed distribution of valuations and thin markets
  - Efficient contracting requires paying large informational rent
- Online ad auctions
  - Conjecture: Narrow targeting can soften the competition and create thin markets [Levin and Milgrom, 2010]
  - More targeting may hurt platform revenues
    - Limited empirical evidence
- What is the optimal level of targeting for the platform?
  - Possibility of self-regulation if the platform limits behavioral targeting

Research Agenda

- Targeting and efficiency
  - How can ad-networks develop targeting policies?
  - How can we evaluate the performance of these policies?
- Value of targeting information
  - How do different pieces of information contribute towards improving targeting ability?
    - Value of contextual (when and where) vs. Behavioral (who)
- Revenue-efficiency trade-off and platform's incentives
  - What is the empirical relationship between efficiency and platform revenues?
  - What is the optimal level of targeting from the perspective of different players?
    - Does the platform have incentives to preserve user privacy?

Key Challenges

- Need counterfactual CTR estimates for ads not shown
  - To develop an efficient targeting policy
- Need a targeting framework with high predictive accuracy
  - To accurately estimate value of each piece of information
- Need an economic model of strategic interactions
  - To provide reliable estimates of market outcomes
  - To examine platform's incentives to preserve users' privacy

Our Approach

- Filtering procedure for counterfactual estimates
  - Identify ads in each impression that could have been shown
- Machine learning CTR prediction model
  - High predictive accuracy
  - Feature generation and categorization to capture and measure the value of each piece of information
- Economic model of auctions to determine market outcomes under counterfactual targeting regimes
  - Characterize advertisers' utility function in targeting scenarios
  - Estimate market outcomes – total surplus and platform revenues
  - Optimal targeting from platform's and advertisers' perspectives

Our Approach

- Filtering procedure for counterfactual estimates
  - Identify ads in each impression that could have been shown
- Machine learning CTR prediction model
  - High predictive accuracy
  - Feature generation and categorization to capture and measure the value of each piece of information
- Economic model of auctions to determine market outcomes under counterfactual targeting regimes
  - Characterize advertisers' utility function in targeting scenarios
  - Estimate market outcomes – total surplus and platform revenues
  - Optimal targeting from platform's and advertisers' perspectives

Coherently combine predictive machine learning models with prescriptive economic models

Related Literature

- CTR estimation and targeting
  - Method
    - Friedman et al. (2000), Friedman (2001), Breiman (2001), Chen and Guesterin (2016)
  - Application
    - McMahan et al. (2013), He et al. (2014), Chapelle et al. (2015)
- Interplay between targeting and privacy
  - Effects of privacy regulation
    - Goldfarb and Tucker (2011a), Johnson (2013), Goldfarb (2014)
  - User behavior
    - Goldfarb and Tucker (2011b), Tucker (2014), Acquisti et al. (2016)
- Revenue efficiency trade-off
  - Theoretical
    - Levin and Milgrom (2010), Bergemann and Bonatti (2011), Celis et al. (2014), Amaldoss et al. (2015), Hummel and McAfee (2016), Sayedi (2018)
  - Empirical
    - Athey and Nekipelov (2010), Yao and Mela (2010)

Setting and Data

Setting

- Major in-app advertising platform
  - 85% market share
  - Over 50 million impressions served daily
- Only one format of ad
  - Small banner ad in jpg or gif format in the bottom
- Limited targeting provision
  - Advertisers can only target on broad categories
    - App Category, Province, Brand, Connectivity, MSP, ISP
    - No behavioral targeting
- Quasi-proportional auction

$$\pi_a = \frac{b_a q_a}{\sum_{j \in \mathcal{A}} b_j q_j}$$

  - Platform does not personalize or update quality scores
  - Probabilistic allocation rule creates randomization

Data

- Impression-level data from Sep 30 to Oct 30, 2015
    - 1,594,831,699 impressions
    - 14,373,29 clicks
    - 0.0090 CTR
- Variables
    - Time and date
    - App ID
    - Device ID
    - Ad ID
    - Targeting variables
        - App Category, Province, Hour, Brand, Connectivity, MSP, ISP
    - Exact location
        - Latitude, Longitude
    - Click indicator

Setting

- Major in-app advertising platform
  - 85% market share
  - Over 50 million impressions served daily
- Only one format of ad
  - Small banner ad in jpg or gif format in the bottom
- Limited targeting provision
  - Advertisers can only target on broad categories
    - App Category, Province, Brand, Connectivity, MSP, ISP
    - No behavioral targeting
- Quasi-proportional auction

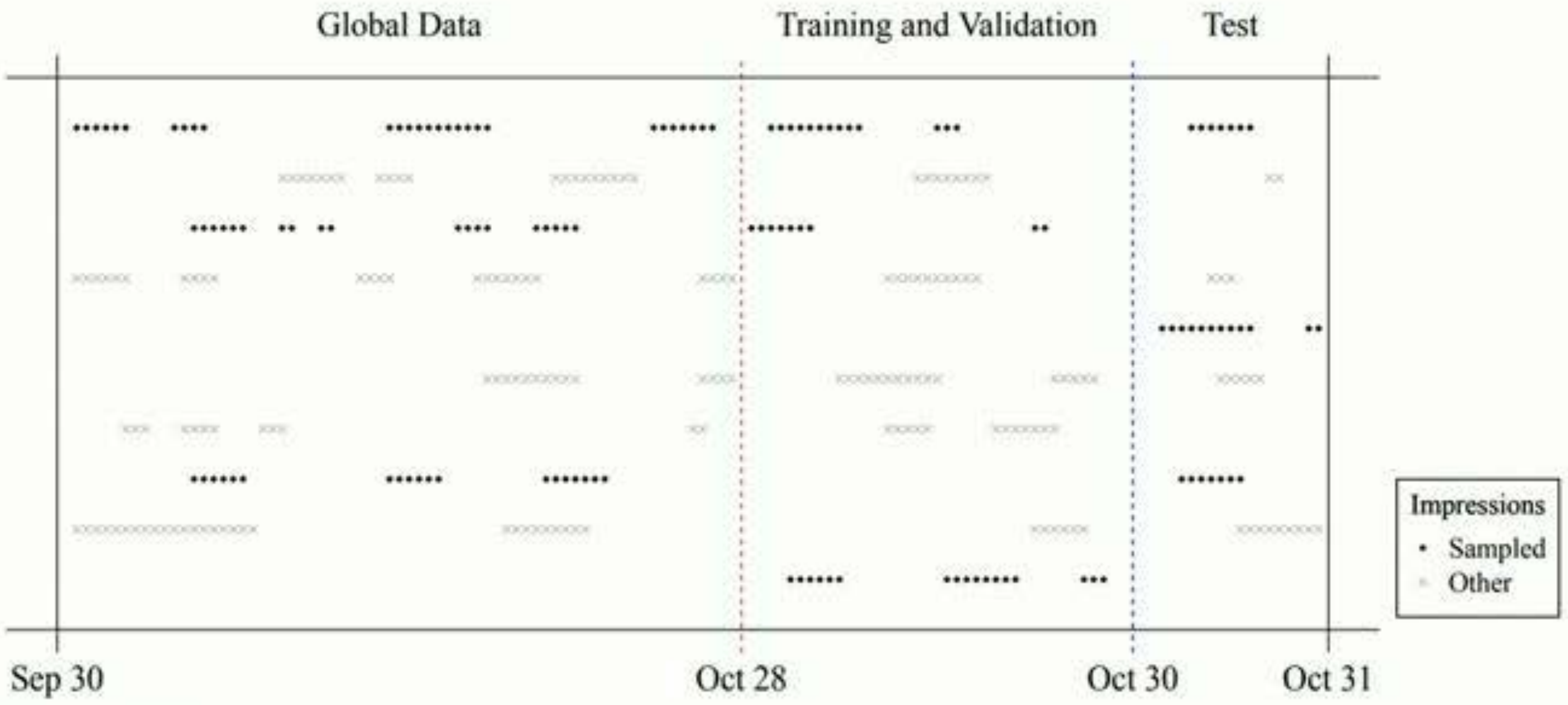
$$\pi_a = \frac{b_a q_a}{\sum_{j \in \mathcal{A}} b_j q_j}$$

  - Platform does not personalize or update quality scores
  - Probabilistic allocation rule creates randomization

Data

- Impression-level data from Sep 30 to Oct 30, 2015
  - 1,594,831,699 impressions
  - 14,373,29 clicks
  - 0.0090 CTR
- Variables
  - Time and date
  - App ID
  - Device ID
  - Ad ID
  - Targeting variables
    - App Category, Province, Hour, Brand, Connectivity, MSP, ISP
  - Exact location
    - Latitude, Longitude
  - Click indicator

Data Splits and Sampling

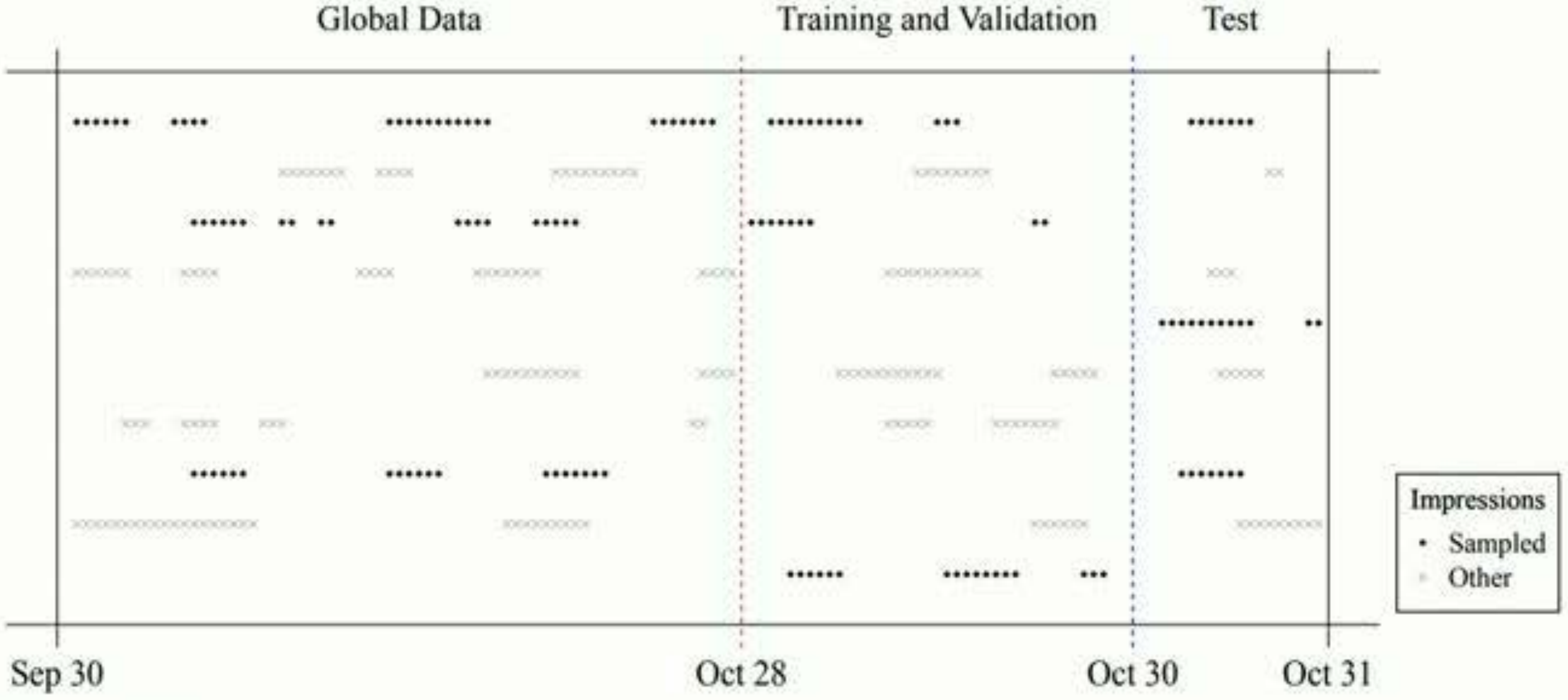


- Data splits
  - Global data (over 146 million impressions)
  - Training, validation and test (over 27 million impressions)
- Sampling procedure
  - Use the full history for sampled users (over 700K users)
    - Data sufficiency is shown for robustness

Data

- Impression-level data from Sep 30 to Oct 30, 2015
  - 1,594,831,699 impressions
  - 14,373,29 clicks
  - 0.0090 CTR
- Variables
  - Time and date
  - App ID
  - Device ID
  - Ad ID
  - Targeting variables
    - App Category, Province, Hour, Brand, Connectivity, MSP, ISP
  - Exact location
    - Latitude, Longitude
  - Click indicator

Data Splits and Sampling



- Data splits
  - Global data (over 146 million impressions)
  - Training, validation and test (over 27 million impressions)
- Sampling procedure
  - Use the full history for sampled users (over 700K users)
    - Data sufficiency is shown for robustness

Part I

Machine Learning Framework for Targeting

Problem Definition

- Problem: How can we estimate the CTR or match value for ad *a* in impression *i*?

$$\text{Impressions}\ \underbrace{\begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,A} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \cdots & m_{N,A} \end{bmatrix}}_{\text{Ads}}, \quad m_{i,a} = \Pr(y_{i,a} = 1)$$

- Goal
  - Accurately estimate elements of the match value matrix
  - Develop targeting policies that map impressions to ads
- Challenges
  - Counterfactual CTR estimation
  - High-dimensional categorical inputs
  - Predictive accuracy

Problem Definition

- Problem: How can we estimate the CTR or match value for ad *a* in impression *i*?

$$\text{Impressions}\ \underbrace{\begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,A} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \cdots & m_{N,A} \end{bmatrix}}_{\text{Ads}}, \quad \boxed{m_{i,a} = \Pr(y_{i,a} = 1)}$$

- Goal
  - Accurately estimate elements of the match value matrix
  - Develop targeting policies that map impressions to ads
- Challenges
  - Counterfactual CTR estimation
  - High-dimensional categorical inputs
  - Predictive accuracy

Counterfactual CTR Estimation

- Accuracy of CTR estimates
  - The same joint distribution of covariates and outcome in training and test data

$$\mathcal{F}_{train}(X,Y) \sim \mathcal{F}_{test}(X,Y)$$

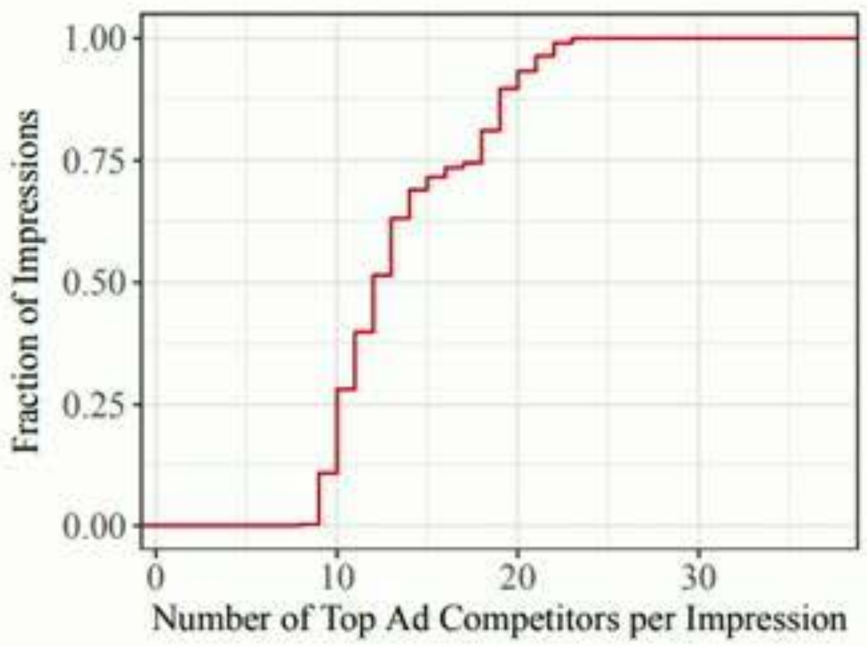
  - Estimates are accurate if the ad *could have been shown* in an impression

$$\Pr(a_i = a) > 0$$

- Key requirement: Randomization
  - Limited targeting provision and no behavioral targeting
    - Ads are shown in a broad set of apps, users, and settings
  - Probabilistic allocation rule
    - Counterfactual CTR estimation fails in second-price auctions

Empirical Strategy

- Filtering procedure
  - Identify availability of an ad for an impression
    - Targeting decision (e.g., excluding a specific province)
    - Campaign availability (e.g., budget exhaustion)
  - Filter ads that are not available
    - Only applicable to impressions with no missing variable (Filtered Sample)
    - Focus on top 37 ads (generate over 80% of total traffic)



Problem Definition

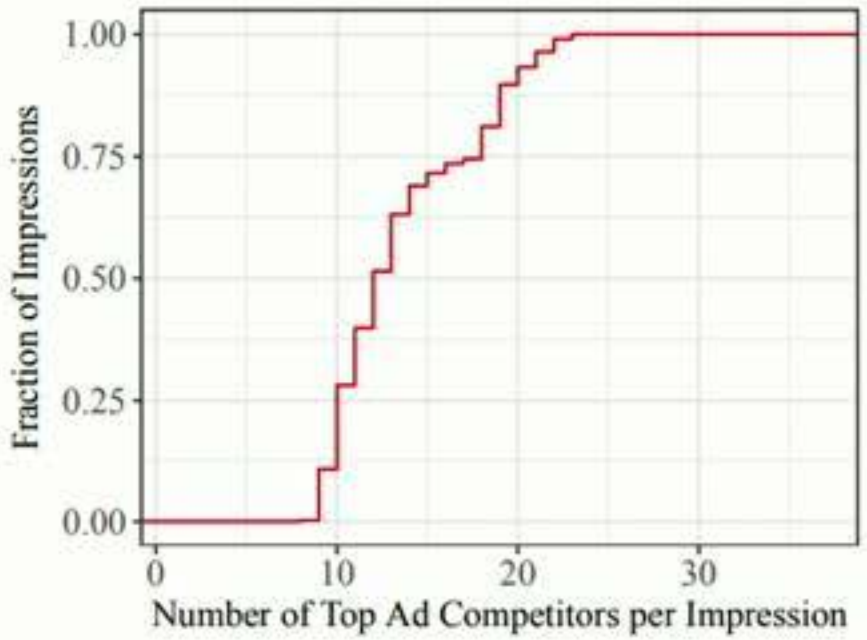
- Problem: How can we estimate the CTR or match value for ad *a* in impression *i*?

$$\text{Impressions}\ \underbrace{\begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,A} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \cdots & m_{N,A} \end{bmatrix}}_{\text{Ads}}, \quad \boxed{m_{i,a} = \Pr(y_{i,a} = 1)}$$

- Goal
  - Accurately estimate elements of the match value matrix
  - Develop targeting policies that map impressions to ads
- Challenges
  - Counterfactual CTR estimation
  - High-dimensional categorical inputs
  - Predictive accuracy

Empirical Strategy

- Filtering procedure
  - Identify availability of an ad for an impression
    - Targeting decision (e.g., excluding a specific province)
    - Campaign availability (e.g., budget exhaustion)
  - Filter ads that are not available
    - Only applicable to impressions with no missing variable (Filtered Sample)
    - Focus on top 37 ads (generate over 80% of total traffic)



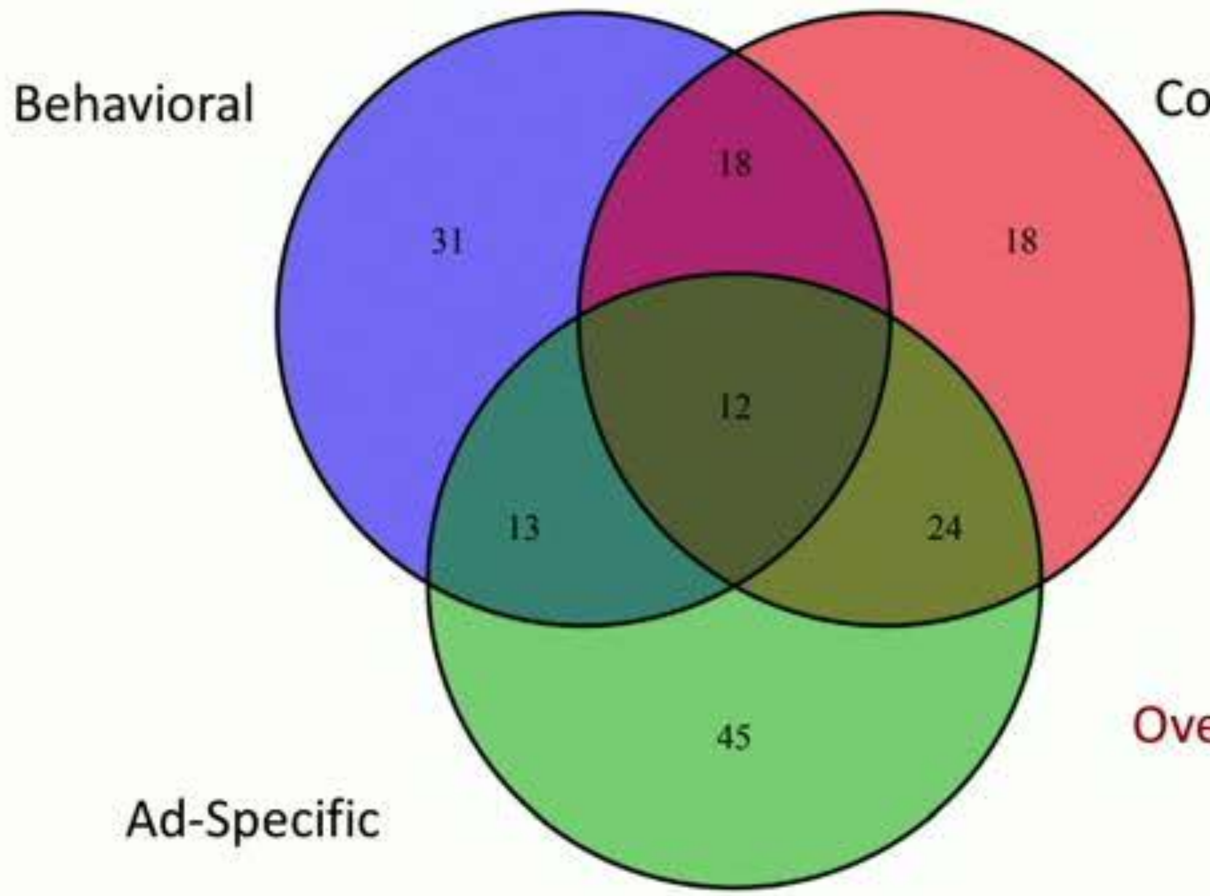
Feature Generation Framework

- Challenge: High dimensional categorical inputs
  - User, App, Ad, Time
- Solution: Functions that map these inputs to meaningful features using the user-level and global history

Feature Generation

- Three types of information
  - Contextual (when and where)
    - App: Gaming app vs. Craigslist app
    - Time of day: At work (10 am) vs. leisure (8 pm)
  - Behavioral (who)
    - Related to the user's past app usage, ad exposure, and ad response
  - Ad-related
    - Captures information on the relative performance of different ads
- Three types of history
  - Long-term (over a one month period)
  - Short-term (within the last week)
  - Ongoing session-level (within this session)

Feature Categorization



Overall 160 features

Learning Algorithm

- Log Loss as objective function

$$\mathcal{L}^{log\ loss}(\hat{\mathbf{M}}, \mathbf{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_{i,a_i}\log\left(\hat{m}_{i,a_i}\right) + (1 - y_{i,a_i})\log\left(1 - \hat{m}_{i,a_i}\right)\right)$$

  - Faster convergence [Rosasco et al., 2004]
  - Most commonly used loss function in CTR prediction [Yi et al., 2013]

Learning Algorithm

- Log Loss as objective function

$$\mathcal{L}^{log\ loss}(\hat{\mathbf{M}}, \mathbf{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left(y_{i,a_i}\log\left(\hat{m}_{i,a_i}\right) + (1 - y_{i,a_i})\log\left(1 - \hat{m}_{i,a_i}\right)\right)$$

  - Faster convergence [Rosasco et al., 2004]
  - Most commonly used loss function in CTR prediction [Yi et al., 2013]
- Validation
  - Hold-out validation set for tuning hyper-parameters
- XGBoost as learning algorithm [Chen and Guesterin, 2016]
  - Fast and scalable version of Boosted Regression Trees
  - Most successful algorithm in Kaggle contests
  - Model comparison for robustness check
    - Least Squares, LASSO, Logistic Regression, Regression Trees, Random Forest

Model Evaluation I

- Relative Information Gain (RIG) as a measure of fit

$$RIG(\hat{\mathbf{M}}, \mathbf{y}) = \left[1 - \frac{\mathcal{L}^{log\ loss}(\hat{\mathbf{M}}, \mathbf{y})}{\mathcal{L}^{log\ loss}(\bar{\mathbf{y}}, \mathbf{y})}\right] \times 100$$

$$\begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,A} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \cdots & m_{N,A} \end{bmatrix}$$

Model Evaluation I

- Relative Information Gain (RIG) as a measure of fit

$$RIG(\hat{\mathbf{M}}, \mathbf{y}) = \left[1 - \frac{\mathcal{L}^{log\ loss}(\hat{\mathbf{M}}, \mathbf{y})}{\mathcal{L}^{log\ loss}(\bar{\mathbf{y}}, \mathbf{y})}\right] \times 100$$

Test data

$$\begin{bmatrix} m_{1,1} & \boxed{m_{1,2}} & \cdots & m_{1,A} \\ m_{2,1} & m_{2,2} & \cdots & \boxed{m_{2,A}} \\ \vdots & \vdots & \ddots & \vdots \\ \boxed{m_{N,1}} & m_{N,2} & \cdots & m_{N,A} \end{bmatrix}$$

Red elements indicate ads which were actually shown

Model Evaluation I

- Relative Information Gain (RIG) as a measure of fit

$$RIG(\hat{\mathbf{M}}, \mathbf{y}) = \left[1 - \frac{\mathcal{L}^{log\ loss}(\hat{\mathbf{M}}, \mathbf{y})}{\mathcal{L}^{log\ loss}(\bar{\mathbf{y}}, \mathbf{y})}\right] \times 100$$

Test data

$$\begin{bmatrix} m_{1,1} & \boxed{m_{1,2}} & \cdots & m_{1,A} \\ m_{2,1} & m_{2,2} & \cdots & \boxed{m_{2,A}} \\ \vdots & \vdots & \ddots & \vdots \\ \boxed{m_{N,1}} & m_{N,2} & \cdots & m_{N,A} \end{bmatrix}$$

Red elements indicate ads which were actually shown

- RIG allows us to evaluate model performance based on actual data
- Consistent with our loss function
- Can be used to quantify the gains from different feature categories
- Other evaluation metrics for robustness check
  - MSE, AUC, 0/1 Loss, Confusion Matrix

Model Evaluation II

- Potential improvement in CTR
  - Based on counterfactual outcomes
  - Efficient targeting policy $\tau^*(i) = \arg\max_a \hat{m}_{i,a}$

Test data

$$\begin{bmatrix} m_{1,1} & \boxed{m_{1,2}} & \cdots & \boxed{m_{1,A}} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \cdots & m_{N,A} \end{bmatrix}$$

- Red indicates actual ad shown
- Green indicates optimal ad based on our model

  - Improvement in CTR using efficient targeting policy

$$\rho(\tau^*, \tau_0; N_F) = \frac{\hat{m}^{\tau^*}}{\hat{m}^{\tau_0}} = \frac{\frac{1}{N_F}\sum_{i=1}^{N_F} \hat{m}_{i,\tau^*(i)}}{\frac{1}{N_F}\sum_{i=1}^{N_F} \hat{m}_{i,\tau_0(i)}}$$

RIG and Value of Information

- Different models
  - Contextual Model: purely contextual + ad-specific features
  - Behavioral Model: purely behavioral + ad-specific features
  - Full Model: all features

RIG and Value of Information

- Different models
  - Contextual Model: purely contextual + ad-specific features
  - Behavioral Model: purely behavioral + ad-specific features
  - Full Model: all features
- Results on predictive accuracy

RIG over Baseline	**Full Sample**	**Filtered Sample**
Behavioral Model	12.14%	14.74%
Contextual Model	5.25%	6.77%
Full Model	17.95%	22.45%
No. of Impressions	9,625,835	4,454,634
% of Test Data	100%	46.28%

RIG and Value of Information

- Different models
  - Contextual Model: purely contextual + ad-specific features
  - Behavioral Model: purely behavioral + ad-specific features
  - Full Model: all features
- Results on predictive accuracy

RIG over Baseline	**Full Sample**	**Filtered Sample**
Behavioral Model	12.14%	14.74%
Contextual Model	5.25%	6.77%
Full Model	17.95%	22.45%
No. of Impressions	9,625,835	4,454,634
% of Test Data	100%	46.28%

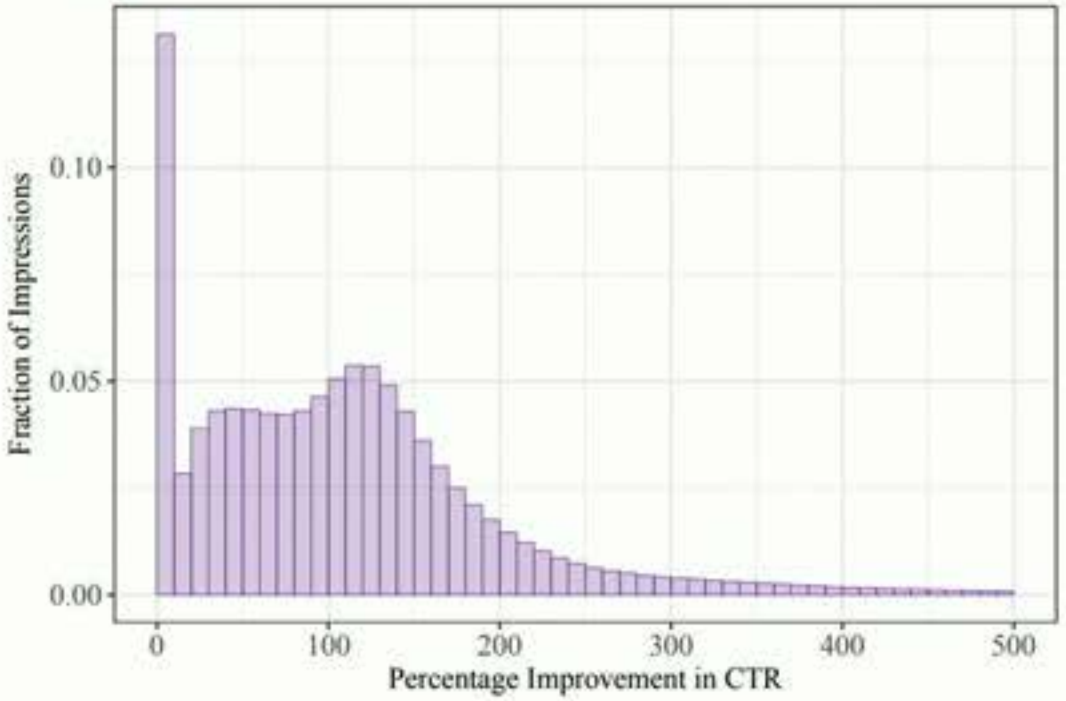
  - Full model reaches 17.95% RIG over the baseline
  - Behavioral information contributes more than contextual
    - While behavioral information impinges on users' privacy, it also significantly improves targeting efficiency

Part II

Revenue-Efficiency Trade-off

Counterfactual CTR Improvement

- Average CTR improvement
  - 66.80% improvement in avg. CTR over the current system
    - Current CTR: 0.66%, Efficient targeting CTR: 1.10%
- Impression-level improvement

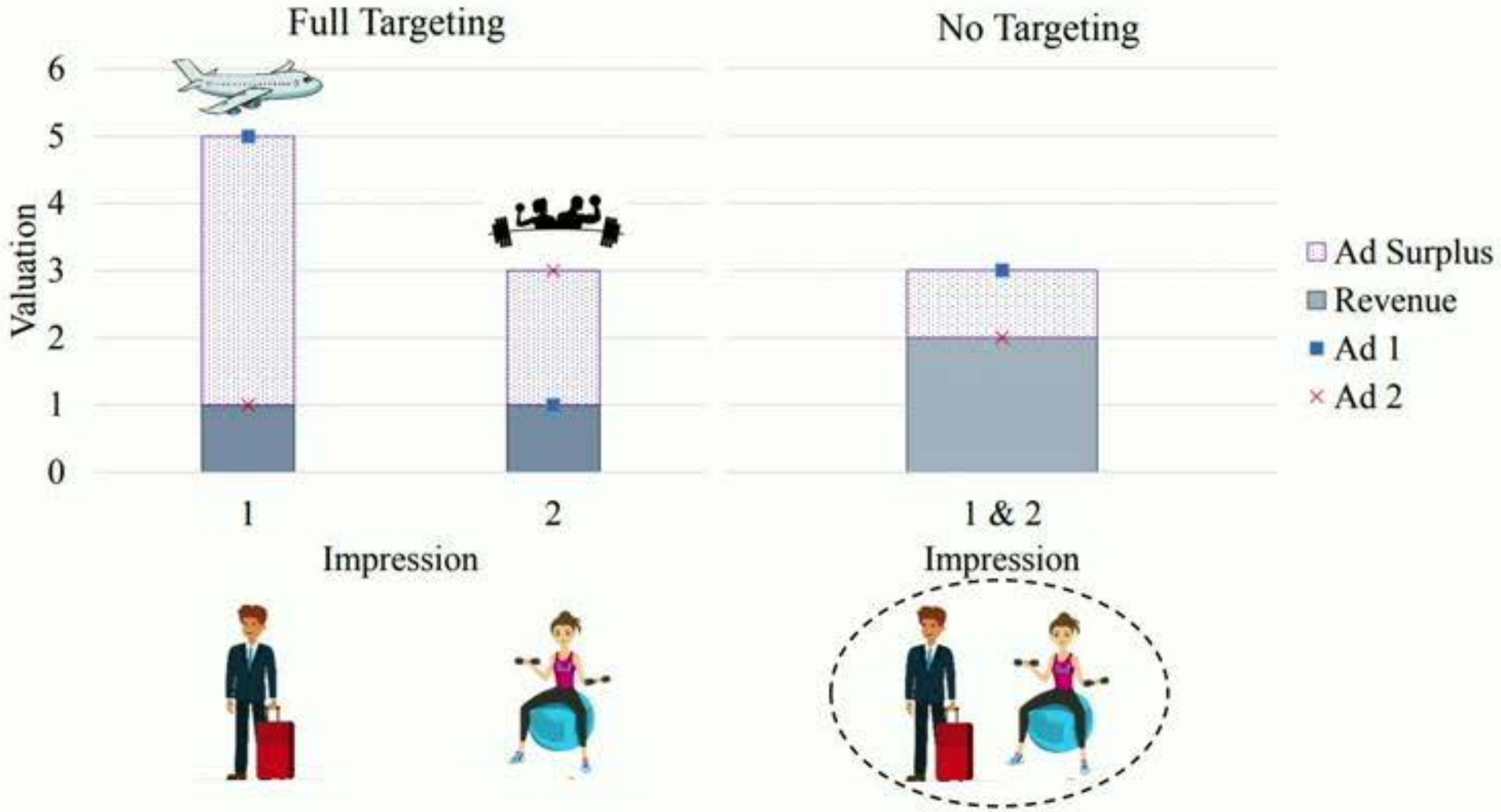


  - Median improvement is 105.35%

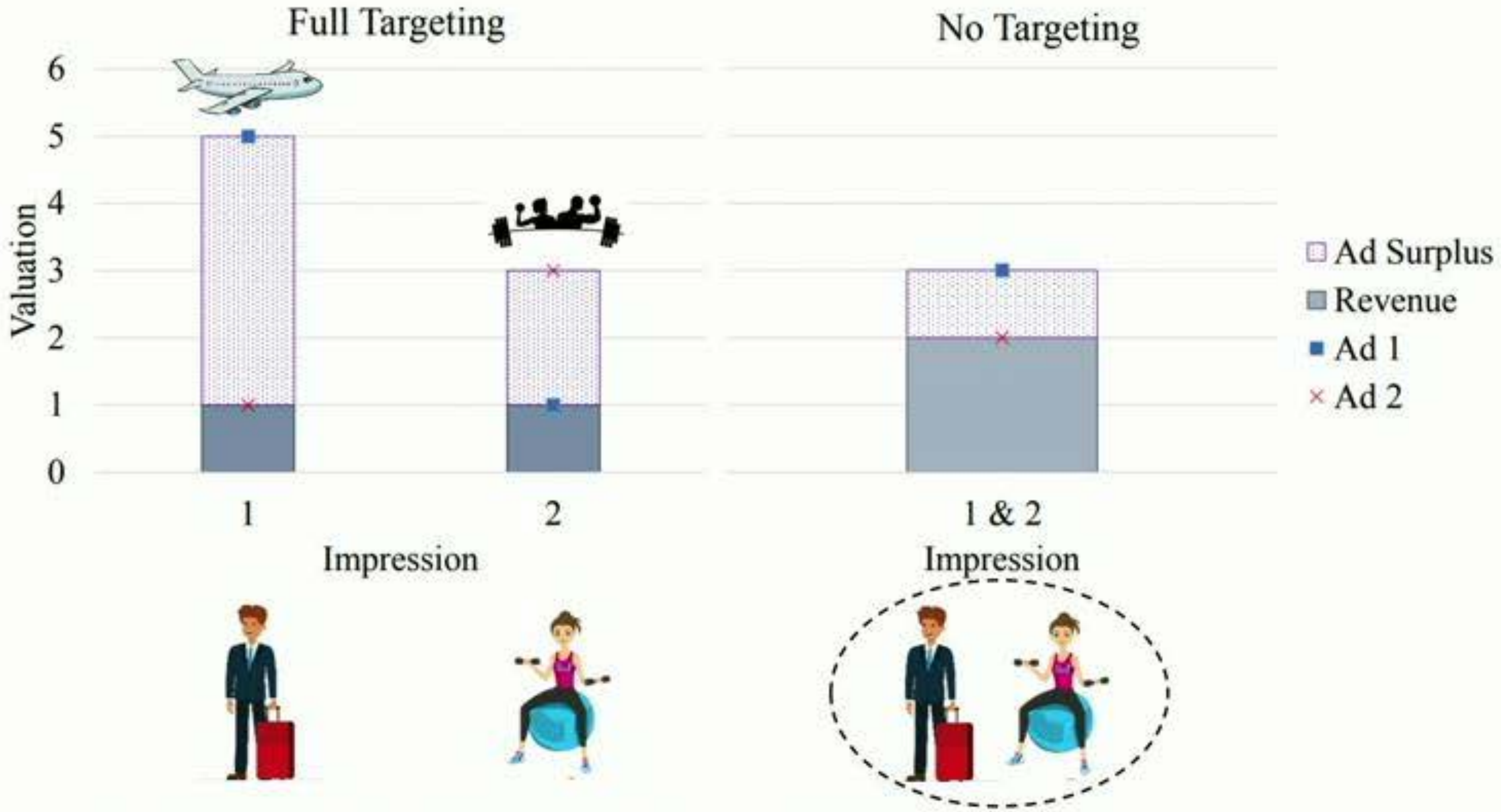
Part II

Revenue-Efficiency Trade-off

Revenue-Efficiency Trade-off



Revenue-Efficiency Trade-off



- More targeting can hurt platform revenues
- What is the optimal level of targeting for the platform?

Model of Auction with Targeting

- Ads' value per impression

$$V = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,A} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ v_{N,1} & v_{N,2} & \cdots & v_{N,A} \end{bmatrix}$$

  - No functional form assumptions on valuations
- Targeting strategy
  - Denotes the platform's decision to bundle N impressions into L bundles such that $\mathcal{I} = \{I_1, I_2, \ldots, I_L\}$
  - Advertiser's valuation for impressions in a bundle: $\frac{1}{|I_j|} \sum_{k \in I_j} v_{ka}$
- Relative granularity of targeting levels
  - Targeting strategy A is at least as granular as B, if two impressions that are distinguishable in B are also distinguishable in A

Analytical Results

- In a second-price auction, as granularity of targeting increases:
  - Total surplus or efficiency increases
  - However, platform revenues can go in either direction
- Four targeting scenarios
  - No targeting: no targeted bidding
  - Contextual targeting: can target at app-time level
  - Behavioral targeting: can target at user-level
  - Full targeting: can target at impression level
- Theoretically:
  - Surplus: $S^{\mathcal{F}} \geq S^{\mathcal{C}}, S^{\mathcal{B}} \geq S^{\mathcal{N}}$
  - Platform revenues: No theoretical guidance

The optimal level of targeting from the platform's perspective is therefore an empirical question

Problem Definition

- Problem: How can we estimate an ad's valuation for each impression under any targeting level?

$$\begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,A} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ v_{N,1} & v_{N,2} & \cdots & v_{N,A} \end{bmatrix}, \quad v_{i,a} = v_a^{(c)} m_{i,a}$$

- Goal
  - Accurately estimate value-per-impression matrix
  - Determine market outcomes – total surplus and platform revenues

- Challenges
  - Estimation of advertisers' click valuations from observed data
  - Estimation of match valuations for any targeting level

Problem Definition

- Problem: How can we estimate an ad's valuation for each impression under any targeting level?

$$\begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,A} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,A} \\ \vdots & \vdots & \ddots & \vdots \\ v_{N,1} & v_{N,2} & \cdots & v_{N,A} \end{bmatrix}, \quad \boxed{v_{i,a} = v_a^{(c)} m_{i,a}}$$

- Goal
  - Accurately estimate value-per-impression matrix
  - Determine market outcomes – total surplus and platform revenues

- Challenges
  - Estimation of advertisers' click valuations from observed data
  - Estimation of match valuations for any targeting level

Estimation of Click Valuations

- Equilibrium for quasi-proportional auction [Mirrokni et al., 2010]

$$\hat{v}_a^{(c)} = \overbrace{b_a^*}^{\text{Equilibrium Bid}} + \frac{b_a^*}{1 - \underbrace{\pi_a}_{\text{Proportion}}}$$

  - If shares are not very high, valuation can be approximated by:

$$\hat{v}_a^{(c)} \approx 2b_a^*$$

  - Unique Bayesian equilibrium in pure strategies when cost function is concave and differentiable
  - Alternative methods for robustness check

Match Valuations Under Targeting

- Define an arbitrary targeting level

$$\mathcal{I} = \{I_1, I_2, \ldots, I_L\}$$

- Aggregation over the bundle
  - Match valuations come from the ML targeting framework's full model

Counterfactual Results

Targeting	Total Surplus	Platform Revenue	Advertisers' Surplus
Full	9.45	8.35	1.10
Behavioral	9.18	8.35	0.84
Contextual	8.99	8.44	0.55
No targeting	8.36	8.30	0.06

- Surplus has a monotonic relationship with granularity
  - Higher efficiency under behavioral targeting compared to contextual targeting
- Revenue is maximized with contextual targeting
  - Platform has natural incentives to limit behavioral targeting

Counterfactual Results

Targeting	Total Surplus	Platform Revenue	Advertisers' Surplus
Full	9.45	8.35	1.10
Behavioral	9.18	8.35	0.84
Contextual	8.99	8.44	0.55
No targeting	8.36	8.30	0.06

- Surplus has a monotonic relationship with granularity
    - Higher efficiency under behavioral targeting compared to contextual targeting
- Revenue is maximized with contextual targeting
    - Platform has natural incentives to limit behavioral targeting
- Privacy implications
    - Self regulation can be achieved!

Conclusion

- Contribution
  - Methodological
    - Scalable machine learning framework for targeting that is compatible with counterfactual analysis of auctions in a competitive environment
  - Substantive
    - Extensive comparison of behavioral and contextual targeting
- Implications
  - Managerial
    - Non-monotonic relationship between revenue and targeting granularity
  - Policy
    - Advertising platforms have incentives to self-regulate

Thank You!