Evaluating User Targeting Policies: Simulation Based on Randomized Experiment Data

WWW2015 FLORENCE, ITALY

24TH INTERNATIONAL WORLD WIDE WEB CONFERENCE
Florence 18 - 22 May 2015

Joel Barajas*[1], Ram Akella[1,2], Marius Holtan[3]
[1]UC Santa Cruz, Santa Cruz CA, USA
[2]School of Information, UC Berkeley, USA
[3]AOL Research, Palo Alto CA, USA

*Contact Author: jbarajas@soe.ucsc.edu, https://users.soe.ucsc.edu/~jbarajas/

University of California Santa Cruz · Aol Advertising. · UC Berkeley School of Information

Abstract

- We propose a user targeting simulator for online display advertising.
- Based on the response of **37 million visiting users** (targeted and non-targeted) and their demographic features, we simulate different user targeting policies.
- We provide evidence that the standard conversion optimization policy shows **similar effectiveness to that of a random targeting**, and significantly inferior to other causally optimized targeting policies.

Introduction

- The use of randomized experiments is becoming the standard practice to accurately measure the ad casual effect on user conversions [2].
  - **Targeted and non-targeted** users convert in the advertiser's website potentially
  - Converting users **regardless of the ad exposure** have motivated the causal analysis of campaign effect
- User targeting development has focused largely on optimizing user conversions by serving ads to the users who **are more likely to convert** [3].
- Often the evaluation of these algorithms is based on **the prediction power of conversions**, which are likely to be not caused by the campaign [2].

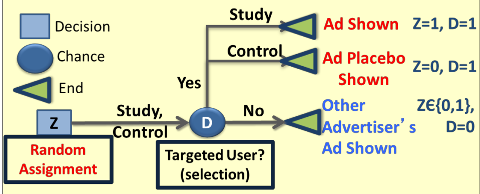
Our Contribution

- We propose a **targeting simulator** that leverages the data of randomized experiments by considering **all the visiting users to the publisher websites** [1].
- We fit the **user conversion response** of the campaign/placebo ad exposures (targeted users), and the response of those who are not targeted.
- Based on the data of a randomized experiment for **37 million users, 8 million targeted users**, and user demographic features, we simulate the standard conversion optimization policy and three targeting algorithms based on the ad average causal effect.

Data

- We consider the user features: **age, gender and income**; segmented by value ranges
- The campaign running time is **two weeks**.
- **Study group**: the total and targeted population sizes are 18.74 and 4.01 million.
- **Control group**: the total and targeted population sizes are 18.70 and 4.09 million.
- **Missing values** are considered as a feature value: **81.4%** of users have one or more missing values.
- We use the first half of the campaign as training, and the second half for testing.

Randomized Experimental Design



- We randomly assign the online **visiting users** to the control or the study treatment arms.
- In practice, a placebo campaign, which **replicates the focal campaign targeting**, is run to display the placebo ads.

User Targeting Simulation

Input/Output

Input: Targeting function $F_{targ}(X_i)$, User Counts $N_z^{obs} = \{N_{dz}^{y}|X_i : \forall d \in \{0,1\}, \forall y \in \{0,1\}, \forall X_i\}$.
Output: Aggregated User Counts After Targeting $N_{z,agg}^{new} = \{N_{dz}^{y,new} : \forall d \in \{0,1\}, \forall y \in \{0,1\}\}$

User Targeting function

Initialization

Set $[\hat{\gamma}_{0z}, \hat{\gamma}_{1z}] = [\mathrm{glmfit}\left([N_{0z}^1|X_i, N_{0z}^0|X_i], \forall X_i\right), \mathrm{glmfit}\left([N_{1z}^1|X_i, N_{1z}^0|X_i], \forall X_i\right)]$ // *Probit Approximation*

Set $[\hat{\theta}_{0z}, \hat{\theta}_{1z}]|X_i = [\Phi(X_i'\hat{\gamma}_{0z}), \Phi(X_i'\hat{\gamma}_{1z})], \quad \forall X_i$ // *Observed Conversion Propensity*

Set $N_z^{Visit}|X_i = N_{1z}^1 + N_{1z}^0 + N_{0z}^1 + N_{0z}^0|X_i, \quad \forall X_i$ // *Audience per Segment X_i*

Set $N_{1z}^{budget} = \sum_{\forall X_i}(N_{1z}^1 + N_{1z}^0)|X_i$ // *Observed Budget*

Set $N_{1z}^{1,new}|X_i = N_{1z}^{0,new}|X_i = 0, \quad \forall X_i$ // *Set Counts*

Set $N_{remain}^{budget} = N_{1z}^{budget}$

Observed targeted users represent a fixed budget.

User Response and Targeting

while $N_{remain}^{budget} > 0$ **do**

Set $P(X_i) = N_{remain}^{Visit}|X_i / \sum_{\forall X_i} N_{remain}^{Visit}|X_i, \quad \forall X_i$

Set $\lambda = N_{remain}^{budget} / \left(\sum_{\forall X_i} N_{remain}^{budget} \times F_{targ}(X_i) \times P(X_i)|X_i\right)$ // *Budget Multiplier*

Set $\left[N_{1z}^{1,new}, N_{1z}^{0,new}\right]|X_i = \left[N_{1z}^{1,new}, N_{1z}^{0,new}\right]|X_i + \min\left(\lambda \times F_{targ}(X_i) \times N_{remain}^{budget} \times P(X_i), N_{remain}^{Visit}|X_i\right) \times \left[\hat{\theta}_{1z}, 1 - \hat{\theta}_{1z}\right]|X_i, \quad \forall X_i$ // *Target Users*

Set $N_{remain}^{Visit}|X_i = N_z^{Visit} - (N_{1z}^{1,new} + N_{1z}^{0,new})|X_i, \quad \forall X_i$ // *Remaining Audience*

Set $N_{remain}^{budget} = N_{1z}^{budget} - \left(\sum_{\forall X_i}[N_{1z}^{1,new} + N_{1z}^{0,new}|X_i]\right)$ // *Remaining Budget*

end while

- **Campaign budget is consumed by the user targeting including the probability of user segments**
- **The visiting population segment constraints is enforced**
- **The while loop re-distributes the remaining budget**

User Counts Aggregation

Set $\left[N_{0z}^{1,new}, N_{0z}^{0,new}\right]|X_i = N_{remain}^{Visit} \times \left[\hat{\theta}_{0z}, 1 - \hat{\theta}_{0z}\right]|X_i, \quad \forall X_i$ // *Non-Targeted User Counts*

Set $N_{z,agg}^{new} = \left\{\sum_{\forall X_i} N_{dz}^{y,new}|X_i : \forall d \in \{0,1\}, \forall y \in \{0,1\}\right\}$ // *Aggregate User Counts*

- **Aggregate the user counts over Xi**
- **This simulation is run for both treatment arms independently**
- **The ad effect is measured based on a t-test**

Validation

- We test the targeting policies with training data:
  1. Random, ***F(Xi) = 1***,
  2. Conversion optimization
  3. Maximization/minimization of ATE, {***ATE(Xi), −ATE(Xi)***}.
  4. ATE maximization, where the segments with negative ATE are set to the minimum positive ATE (***ATE+(Xi)***),
  5. ATE minimization of ATE (***−ATE−(Xi)***).

Table 1: Simulator Validation. Targeting functions are trained and tested with the same data. ATE intervals are shown for 0.10 significance level.

$F_{targ}(X_i)$	ATE (1e-6)	*lift* (%)	$F_{targ}(X_i)$	ATE (1e-6)	*lift* (%)
1(Random)	3.76±9.83	**7.37**	$\theta_{11}\|X_i$	2.92±10.0	5.46
ATE(X_i)	5.63±9.62	11.77	-ATE(X_i)	-1.74±10.3	-2.94
ATE$^+$(X_i)	8.7[illegible]	**19.26**	-ATE$^-$(X_i)	-6.68[illegible]	**-9.78**

- **Maximizing ATE shows the best performance, and minimizing ATE the worst perfromance.**
- **Both effects are far from the random targeting**

Results

Table 2: Targeting Policy Testing Results. ATE intervals are shown for 0.10 significance level.

	All Users		No Missing Features	
$F_{targ}(X_i)$	ATE(1e-5)	*lift*(%)	ATE(1e-5)	*lift*(%)
1 (Random)	1.35±1.74	11.01	2.21±4.26	14.06
$\theta_{11}/(1-\theta_{11})\|X_i$	1.38±1.77	10.91	1.98±3.85	12.25
ATE(X_i)	1.45±1.73	12.00	2.45±4.39	16.25
ATE$^+$(X_i)	1.69±1.76	**13.72**	2.92±3.55	**19.92**
$lift^+(X_i)$	**1.78±1.7[illegible]**	**[illegible]4.47**	3.00[illegible]	**20.87**

- **The user conversion optimization performance is similar to a random targeting**
- **Optimizing the lift shows the best causal attribution performance**
- **The effect results estimated for users with no missing features depict the same directional results**

Conclusion and Discussion

- We have found evidence that the standard **practice of optimizing the conversion probability does not optimize the causal effect of the ad.**
- We have shown that the user targeting **makes a difference** in the ad evaluation even when a placebo ad is displayed.
- This finding **contradicts the standard evaluation practice** of measuring the effect with a non-optimized campaign, which is assumed to hold for future optimized exposures.

Acknowledgements

- This work is partially funded by CONACYT UC-MEXUS grant 194880, CITRIS and AOL Faculty Award.

References

[1] J. Barajas, J. Kwon, R. Akella, A. Flores, M. Holtan, and V. Andrei. Marketing campaign evaluation in targeted display advertising. In ADKDD '12, pages 1–7. ACM, 2012.
[2] R. A. Lewis, J. M. Rao, and D. H. Reiley. Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of WWW2011*, pages 157–166. ACM, 2011.
[3] S. Pandey, M. Aly, A. Bagherjeiran, A. Hatch, P. Ciccolo, A. Ratnaparkhi, and M. Zinkevich. Learning to target: What works for behavioral targeting. In CIKM '11, pages 1805–1814, 2011.