Matching Ads to Users: Performance analysis of an Advertising Search Engine

Miguel Alcobendas, Naji Shajarisales

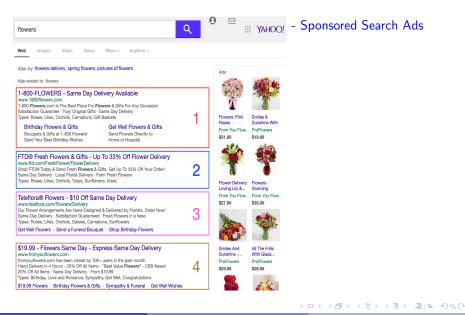
Yahoo! Research - Carnegie Mellon University

April 27, 2019

Miguel Alcobendas

EL SQA

Introduction:

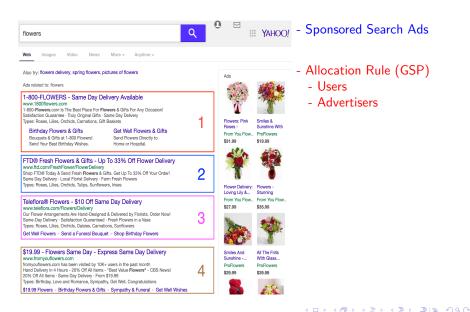


Miguel Alcobendas

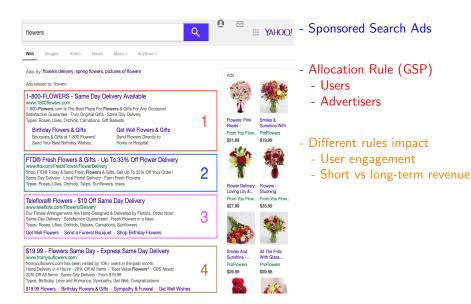
Ads to Users

April 27, 2019 2 / 50

Introduction:



Introduction:



Miguel Alcobendas

Ads to Users

▲□▶ ▲□▶ ▲ヨ▶ ▲ヨ▶ ヨヨ ののべ

Introduction

• Objectives & Contributions:

- Study the bias of Yahoo's search engine towards (at scale):
 - Allocation maximizing users' welfare
 - Allocation maximizing value-per-click of advertisers
- Relationship between users' perceived ad quality and value-per-click of advertisers

ELE SQC

• • = • • =

Introduction

• Objectives & Contributions:

- Study the bias of Yahoo's search engine towards (at scale):
 - Allocation maximizing users' welfare
 - Allocation maximizing value-per-click of advertisers
- Relationship between users' perceived ad quality and value-per-click of advertisers
- Problem: Marketplace researchers do not observe:
 - Long-term impact of different allocation rules (tempting rules)
 - Value-per-click of advertisers (allocation rule is not truthful)
 - Perceived quality of ads by users (ad-quality and position effect)

SIN NOR

A B F A B F

Introduction

• Objectives & Contributions:

- Study the bias of Yahoo's search engine towards (at scale):
 - Allocation maximizing users' welfare
 - Allocation maximizing value-per-click of advertisers
- Relationship between users' perceived ad quality and value-per-click of advertisers
- Problem: Marketplace researchers do not observe:
 - Long-term impact of different allocation rules (tempting rules)
 - Value-per-click of advertisers (allocation rule is not truthful)
 - Perceived quality of ads by users (ad-quality and position effect)

• Data:

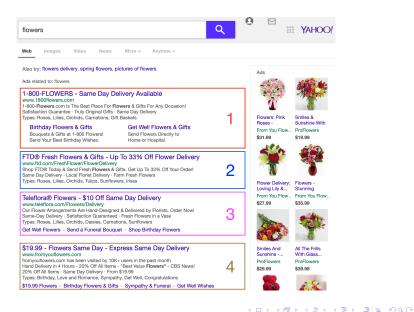
- Yahoo's search engine (2017)
 - Production logs
 - Experiment (Identification)
- Displayed Queries: "cheap flight","map","game","United Airlines","100 Popular Keywords"

Rest of the Presentation

• Allocation & Pricing Rule: Generalized Second Price Auction

- Inference of Value-per-Click of Advertisers
- Inference of Perceived Ad-Quality by Users
- How to compare sequences of ads: Optimal Matching
- Results
- Conclusions

SIN NOR



Allocation Mechanism

- \mathcal{J} advertising slots in the search results page (5 slots in desktop max)
- \mathcal{I} advertisers $> \mathcal{J}$
- Each $i \in \mathcal{I}$ places a per click bid b_i on a single search Keyword
- Yahoo's search engine assigns a scoring factor to advertisers (s_i)
 - Click probability of ad *i* appearing in position 1 ($s_i = Pr_{i1}$)
 - Squashing factor $(s_i = Pr_{i1}^{\theta} \text{ where } \theta >= 0)$
 - Coarsening less accurate estimator for "clickability" (e.g. no gender) different scoring rule may impact revenue and/or ranking
- Advertisers are ranked by their rank-score q_i

$$q_{i} = \underbrace{b_{i}}_{Advertiser} \times \underbrace{s_{i}}_{f(Clickability)}$$

Generalized Second Price Auction Pricing Rule

- Advertiser *i* in position *j* only pays if user clicks
- Payment: Cost-per-click is the minimal bid *i* has to place to keep its position

$$c_{ij}(b) = rac{s_{\pi(j+1)} \cdot b_{\pi(j+1)}}{s_i} \mathbb{1}[j \in \mathcal{J}]^{-1}$$

where $\pi(j+1)$ denotes the advertiser that is located in position j+1

Example 1

	flowers	₩ ҮАНОО!
	Web Images Video News More - Anytime -	
$q_i = s_i imes b_i$	Also try: flowers delivery, spring flowers, pictures of flowers	Ads
	Ads related to: flowers	
$q_a = 0.07 imes 1$	1.800-FLOWERS - Same Day Delivery Available www.1800flowers.com 1.600-Flowers.com <	Flowers: Pink Smiles & Roses - From You Flow, ProFlowers \$31.99 \$19.99
	Send Your Best Birthday Wishes. Home or Hospital.	331.99 \$19.99
$q_b = 0.05 imes 1$	FTD® Fresh Flowers & Gifts - Up To 33% Off Flower Delivery www.ftd.com/FreshFlower/Flower/Plower Shop FTD® Toda & Send Freih Flowers & Gifts. Get Up To 33% Off Your Ordert Same Day Delivery - Local Front Delivery - Farm Freih Flowers Types: Roses, Lilles, Orchids, Tulips, Suntlowers, Irrises	Flower Delivery: Flowers - Loving Lily & Stunning
$q_c = 0.025 imes 1$	Teleflora@ Flowers - \$10 Off Same Day Delivery www.teleflora.com/Flowers/Delivery Our Flower Arrangements Are Hand-Designed & Delivered by Florists. Order Nowl Same-Day Delivery - Statistaction Guaranteed - Fresh Flowers In a Vase Types: Rose, Liles, Orchica, Duelise, Cannations, Wurthowers Get Weil Flowers - Send a Funeral Bouquet - Shop Birthday Flowers	From You Flow \$27.99 \$35.99 \$35.99
$q_d = 0.01 imes 1$	\$19.99 - Flowers Same Day - Express Same Day Delivery www.fromyouflowers.com fromyouflowers.com has been visited by 10K+ users in the past month Hand Delivery in 4 Hours - 20% Off Al Items - Baset Value Flowers ²⁺ -CBS Newel 20% Off Al Items - Same Day Delivery - From S18 Velay Types: Birthday, Love and Romance, Sympathy, Guet Vela, Congratulations \$19.99 Flowers - Birthday Howers A Giffs - Sympathy & Funder - Get Well Wishes	Smiles And Sunshine All The Frills With Glass ProFlowers \$29.99 \$39.98
$q_e = 0.001 imes 1$		

三日 のへで

イロト イヨト イヨト イヨト

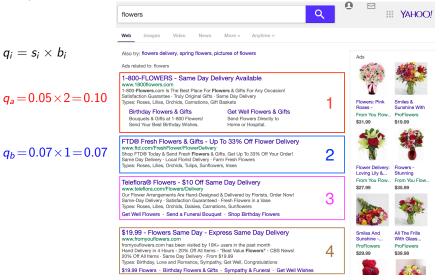
Example 1

n Q YAHOO! flowers Web $q_i = s_i \times b_i$ Also try: flowers delivery, spring flowers, pictures of flowers Ads Ads related to: flowers 1-800-FLOWERS - Same Day Delivery Available www.1800flowers.com $a_{2} = 0.07 \times 1$ 1-800-Flowers.com Is The Best Place For Flowers & Gifts For Any Occasion Satisfaction Guarantee · Truly Original Gifts · Same Day Delivery Types: Roses, Lilies, Orchids, Carnations, Gift Baskets Flowers: Pink Smiles & $c_{a,1} = \frac{0.05 \times 1}{0.07} = 0.7$ Roses -Sunshine With Birthday Flowers & Gifts Get Well Flowers & Gifts From You Flow. ProFlowers Bouquets & Gifts at 1-800 Flowers! Send Flowers Directly to \$31.99 \$19.99 Home or Hospital Send Your Best Birthday Wishes. $q_{b} = 0.05 \times 1$ FTD® Fresh Flowers & Gifts - Up To 33% Off Flower Delivery www.ftd.com/FreshFlower/FlowerDelivery 2 Shop FTD® Today & Send Fresh Flowers & Gifts. Get Up To 33% Off Your Order! $c_{b,2} = \frac{0.025 \times 1}{0.05} = 0.5$ Same Day Delivery - Local Florist Delivery - Farm Fresh Flowers Types: Roses, Lilles, Orchids, Tulips, Sunflowers, Irises Flower Delivery: Flowers -Loving Lily &... Stunning From You Flow., From You Flow., Teleflora® Flowers - \$10 Off Same Day Delivery $a_{c} = 0.025 \times 1$ \$27.99 \$35.99 www.teleflora.com/Flowers/Delivery Our Flower Arrangements Are Hand-Designed & Delivered by Florists. Order Now! Same-Day Delivery · Satisfaction Guaranteed · Fresh Flowers in a Vase $c_{c,3} = \frac{0.01 \times 1}{0.025} = 0.4$ Types: Roses, Lilies, Orchids, Daisies, Carnations, Sunflowers Get Well Flowers - Send a Funeral Bouquet - Shop Birthday Flowers \$19.99 - Flowers Same Day - Express Same Day Delivery All The Frills Smiles And www.fromyouflowers.com Sunshine -... With Glass... $q_d = 0.01 \times 1$ fromyouflowers.com has been visited by 10K+ users in the past month ProFlowers ProFlowers Hand Delivery in 4 Hours - 20% Off All Items - "Best Value Flowers" - CBS News! Δ \$29.99 \$39.98 20% Off All Items · Same Day Delivery · From \$19.99 $c_{d,4} = \frac{0.001 \times 1}{0.01} = 0.1$ Types: Birthday, Love and Romance, Sympathy, Get Well, Congratulations \$19.99 Flowers - Birthday Flowers & Gifts - Sympathy & Funeral - Get Well Wishes

 $q_e = 0.001 \times 1$

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

Example 2



Miguel Alcobendas

April 27, 2019 10 / 50

<□> <同> <同> < 回> < 回> < 回> < 回> < 回> < 回< の< ○

• Truth-telling is **NOT** a dominant strategy for advertisers

JI NOR

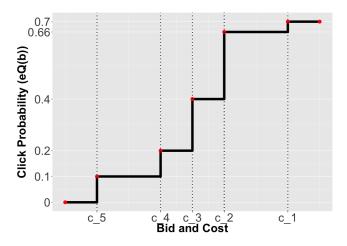
- Truth-telling is **NOT** a dominant strategy for advertisers
- Expected profit of bidder i

$$\pi_i(b; v_i) = (v_i - c_{i\sigma_i(b)}(b))eQ_i(b)$$

where

- v_i: Value-per-click of advertiser i (Unobserved)
- c_{iσi(b)}(b): cost at position σ_i(b)
- $eQ_i(b)$: Allocation probability

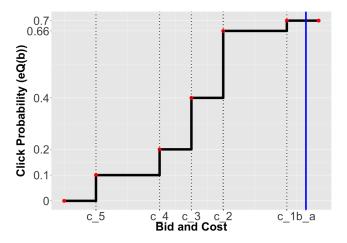
Truth-telling is NOT a dominant strategy



 $c_1 >> c_2 \quad 0.7 \approx 0.66$

-

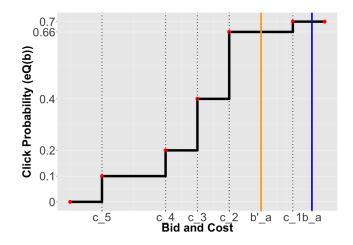
Truth-telling is NOT a dominant strategy



 $\pi_a(b_a; v_a) = (v_a - c_1(b))0.7$

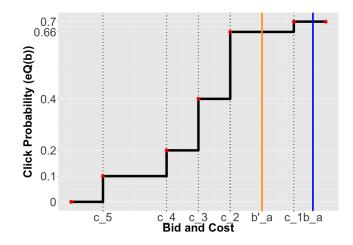
-

Truth-telling is NOT a dominant strategy



 $\pi_{a}(b_{a}^{'};v_{a}) = (v_{a} - c_{2}(b_{a}^{'}))0.66$

Truth-telling is NOT a dominant strategy



 $\pi_a(b_a^{'}; v_a) = (v_a - c_2(b_a^{'}))0.66 > \pi_a(b_a; v_a) = (v_a - c_1(b_a))0.7$

- Allocation & Pricing Rule: Generalized Second Price Auction
- Inference of Value-per-Click of Advertisers
- Inference of Perceived Ad-Quality by Users
- How to compare sequences of ads: Optimal Matching
- Results
- Conclusions

ELE SOC

() < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < ()

- Truth-telling is **NOT** a dominant strategy \Rightarrow Infer Value-per-click
- Infer advertisers' value-per-click using the concept of no-regret learning (Nekipelov, Syrgkanis, and Tardos (2015))
 - Experimenting with bids as off-equilibrium (market exploration)
 - Weaker assumptions than Nash equilibrium
 - Assume that bidders are learning over time the strategy that maximizes their expected profits

• We can express the expected utility of bidder *i* at *t* as

$$\pi_{it}(b_{it}, \mathbf{v}_{i}) = \mathbf{v}_{i} \underbrace{eQ_{it}(b_{it})}_{\text{Click Probability}} - \underbrace{eC_{it}(b_{it})}_{\text{ExpectedCost}}$$

where at each time t bidder i bids b_{it}

ELE NOR

• We can express the expected utility of bidder *i* at *t* as

$$\pi_{it}(b_{it}, \mathbf{v}_{i}) = \mathbf{v}_{i} \underbrace{eQ_{it}(b_{it})}_{\text{Click Probability}} - \underbrace{eC_{it}(b_{it})}_{\text{ExpectedCost}}$$

where at each time t bidder i bids b_{it}

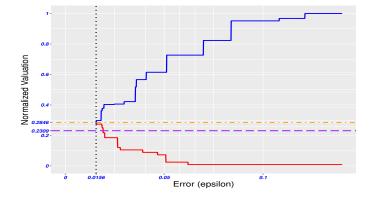
• Rationalizable Set: A pair (ϵ_i, v_i) of value v_i an error ϵ_i is a rationalizable pair for player *i* if it satisfies the below equation.

$$orall b' \in \mathcal{B}: rac{1}{T}\sum_{t=1}^T \pi_{it}(b_{it},v_i) \geq rac{1}{T}\sum_{t=1}^T \pi_{it}(b',v_i) - \epsilon_i$$

where b' corresponds to a fixed bid from the bid sequence $\{b_{it}\}_{t=1}^{T}$ Nash Equilibrium Condition $\rightarrow \epsilon_i = 0 \ \forall i$

• For any error ϵ , the set of values that belongs to the rationalizable set is

$$v_i \in \left[\max_{b':\Delta eQ(b'_i)<0} \frac{eC_{it}(b'_i) - eC_{it}(b_{it}) + \epsilon}{eQ_{it}(b'_i) - eQ_{it}(b_{it})}, \min_{b':\Delta eQ(b'_i)>0} \frac{eC_{it}(b'_i) - eC_{it}(b_{it}) + \epsilon}{eQ_{it}(b'_i) - eQ_{it}(b_{it})}\right]$$



Value-per-click (v_i): the smallest rationalizable error
 ϵ can be viewed as a
 point prediction for the value of the advertiser
 ϵ □ → *ϵ* ⊕ →
 ϵ ⊕ → *ϵ* ⊕ → *ϵ* ⊕ →
 ϵ ⊕ → *ϵ* ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕ →
 ϵ ⊕

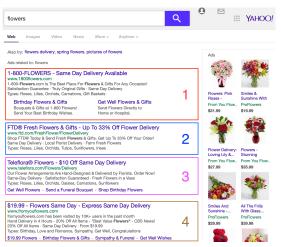
Miguel Alcobenda	35
------------------	----

- Allocation & Pricing Rule: Generalized Second Price Auction
- Inference of Value-per-Click of Advertisers
- Inference of Perceived Ad-Quality by Users
- How to compare sequences of ads: Optimal Matching
- Results
- Conclusions

ELE SQC

() < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < () < ()

Users' Perceived Ad-Quality



- Disentangle ad effect from position effect
- The model accommodates multiple clicks and unobserved correlation across ads within the impression (Jiziorski and Segal (2015)).

Miguel Alcobendas

April 27, 2019 20 / 50

Users' Perceived Ad-Quality

- Utility-maximizing model (allow counterfactuals)
- Assuming no-unobserved correlation among ads, the user's value of clicking on ad *a* located in position *n*

$$\mathcal{U}_{ian} = u_a(x) - f_n + \epsilon_i$$

where

- $u_a(x)$: Perceived utility of ad *a* with features *x*
- f_n : Cost of position n (e.g. scrolling effort, attention)
- ϵ_i : idiosyncratic shock
- Objective: Infer $u_{a(n)}(x)$ and $f_n \forall n, a$
 - $\epsilon \sim$ EV Type 1 Distribution Logistic Model (Maximum likelihood)
 - $\epsilon \sim$ Normal Distribution OLS
- Identification: Randomization of displayed ads

Users' Perceived Ad-Quality

• Assuming unobserved correlation among ads, the user's value of engaging with an impression

$$U_i(\mathcal{C}) = \left(\sum_{n \in \mathcal{C}} \left(u_{a(n)}(x)\right)^{1+R}\right)^{1/(1+R)} - \sum_{n \in \mathcal{C}} f_n + \tau_i$$

where

- $\mathcal{C}:$ set of clicked positions in impression
- $u_{a(n)}(x)$: Perceived utility of ad a displayed in n with features x
- f_n : Cost of position n (e.g. scrolling effort, attention)
- $R\in(-1,\infty)$ captures unobserved correlation among ads in a keyword
- τ_i : idiosyncratic shock
- Objective: Estimate $u_{a(n)}(x)$, f_n and $R \forall n, a$
- Formulate problem as a finite-horizon Markov Decision Process
 - State Variables: displayed ads in an impression
 - Actions: clicks on ads, click on organic links, or conclude search

▲□▶ ▲□▶ ▲ヨ▶ ▲ヨ▶ ヨヨ ののべ

- Allocation & Pricing Rule: Generalized Second Price Auction
- Inference of Value-per-Click of Advertisers
- Inference of Perceived Ad-Quality by Users
- How to compare sequences of ads: Optimal Matching
- Results
- Conclusions

- How to compare sequences of ads: sponsored, rankings maximizing users' welfare and advertisers' value-per-click (Optimal Matching)
- The optimal matching distance equals the smallest number of operations needed to turn one sequence into another
- Operations: insertion, deletion, and substitution (cost)
- The higher the distance, the more dissimilar the ad sequences
- Challenge: size asymmetry of sequences

• Example: Compare impressions with users' optimal sequence

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	cnn.com
Position 2	avf.com
Position 3	pow.com
Position 4	fij.com
Position 5	isw.com
Position 6	kks.com

EL SQA

 In users' sequence, discard non-displayed ads with ranking below the worse ad in the impression

Table: Impression Ads	Tab	ession A	e:	Гаb	٦
-----------------------	-----	----------	----	-----	---

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	cnn.com
Position 2	avf.com
Position 3	pow.com
Position 4	fij.com
Position 5	isw.com
Position 6	kks.com

- ∢ ∃ ▶

EL SQA

 In users' sequence, discard non-displayed ads with ranking below the worse ad in the impression (we do not care about the tail)

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	cnn.com
Position 2	avf.com
Position 3	pow.com
Position 4	fij.com
Position 5	isw.com

11 DQA

- Transformation 1 Deletion. Remove non-displayed ads
- Edit cost: +1 +1 = 2

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	cnn.com
Position 2	avf.com
Position 3	pow.com
Position 4	fij.com
Position 5	isw.com

- Transformation 1 Deletion. Remove non-displayed ads
- Edit cost: +1 +1 = 2

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	avf.com
Position 2	pow.com
Position 3	isw.com

- Transformation 2 Substitution. isw.com by pow.com
- Edit cost: 2 + 2 = 4

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	avf.com
Position 2	pow.com
Position 3	isw.com

A (10) × A (10) × A (10)

ELE SQC

- Transformation 2 Substitution. isw.com by pow.com
- Edit cost: 2 + 2 = 4

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	avf.com
Position 2	pow.com
Position 3	pow.com

- Transformation 3 Substitution. pow.com by isw.com
- Edit cost: 4 + 2 = 6

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	avf.com
Position 2	pow.com
Position 3	pow.com

- Transformation 3 Substitution. pow.com by isw.com
- Edit cost: 4 + 2 = 6

Table: Impression Ads

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

Table: Users' Optimal Sequence

Position 1	avf.com
Position 2	isw.com
Position 3	pow.com

- Allocation & Pricing Rule: Generalized Second Price Auction
- Inference of Value-per-Click of Advertisers
- Inference of Perceived Ad-Quality by Users
- How to compare sequences of ads: Optimal Matching

Results

Conclusions

Data

- Canonized Keywords: "game", "cheap flight", "united airlines", "ups", "map", "100 keywords"
- Individual impression logs from Yahoo's desktop searches collected during the second semester of 2017
- Supplement Yahoo's data with traffic information of ad domains using information from "Alexa.com"
- Users' utility function: experiment where the order of ads is randomized (Identification disentangle ad-quality from position effects)
- Advertisers' value-per-click: production logs (all traffic)

A 回 > A E > A E > A E > A E > A E >

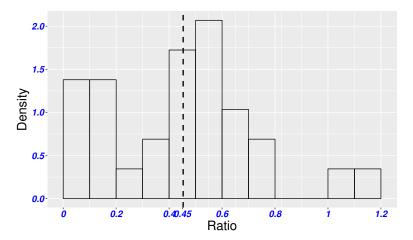


Figure: Game: Average bid to predicted value ratio

Mi	gue	A	lco	bend	las

April 27, 2019 36 / 50

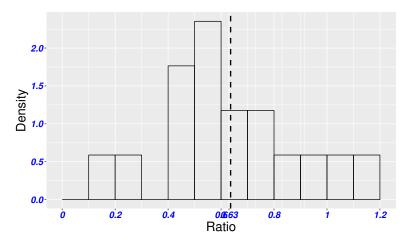


Figure: Cheap flight: Average bid to predicted value ratio

Mi	guel	AI	col	henc	las

April 27, 2019 37 / 50

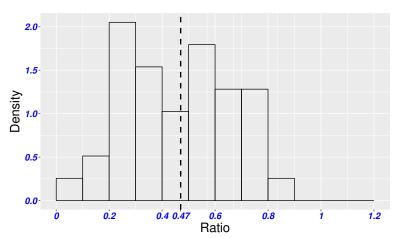
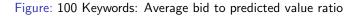


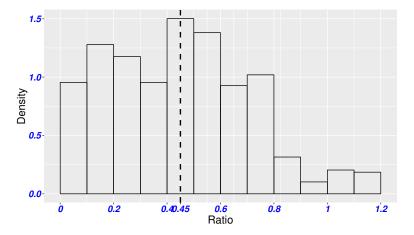
Figure: United Airlines: Average bid to predicted value ratio

Mi	guel	coł	hend	las

April 27, 2019 38 / 50

-





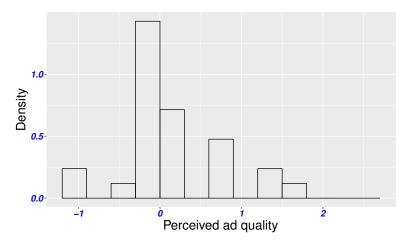


Figure: Game: Ads Utility Distribution

April 27, 2019 40 / 50

三日 のへの

▲ □ ▶ ▲ □ ▶ ▲ □

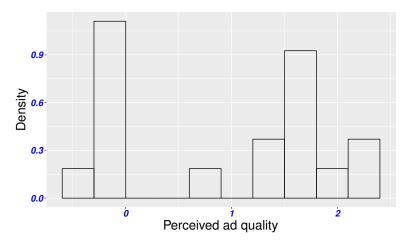


Figure: Cheap Flight: Ads Utility Distribution

Migue	Alco	hend	las

April 27, 2019 41 / 50

三日 のへで

▲圖 ▶ ▲ 臣 ▶ ▲ 臣

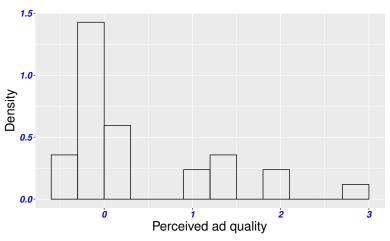


Figure: United Airlines: Ads Utility Distribution

Migue	benc	las

April 27, 2019 42 / 50

三日 のへの

• • = • • =

< 行

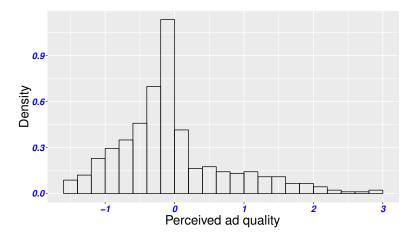


Figure: 100 Keywords: Ads Utility Distribution

April 27, 2019 43 / 50

- 4 ∃ ▶

ELE DOG

• Question: Is there any relationship between users' perceived quality and advertisers' value-per-click?

Answer: Hard to tell

- 4 ∃ ▶

▲ ∃ ► ∃ =

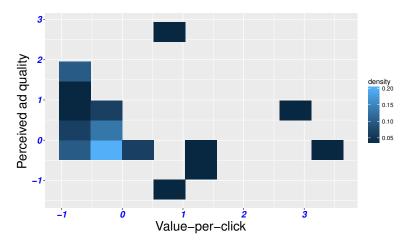


Figure: Game: Ad-quality vs Advertiser's value-per-click

< 1 k

ELE DOG

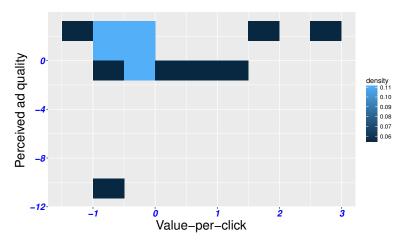


Figure: Cheap flight: Ad-quality vs Advertiser's value-per-click

Migue		

April 27, 2019 46 / 50

- E

ELE DOG

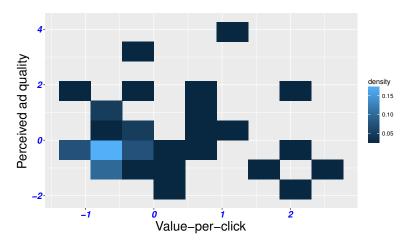


Figure: United Airlines: Ad-quality vs Advertiser's value-per-click

EL SQA

・ 同 ト ・ ヨ ト ・ ヨ

Results: Optimal Matching

• Question: Quantify the bias of Yahoo's search engine

EL SQA

• • • • • • • • • •

Results: Optimal Matching

• Question: Quantify the bias of Yahoo's search engine

	100 KW	Cheap Flight	Game	Мар	United	Ups
Distance	8.88	8.72	12.32	22.57	25.10	9.35
(User)	(4.65)	(4.11)	(8.37)	(5.37)	(8.27)	(6.50)
Distance	10.17	10.06	15.69	16.46	26.96	9.66
(Value-per	(4.90)	(4.78)	(8.83)	(3.88)	(7.38)	(4.49)
-click)						
Distance	-1.29**	-1.34**	-3.36 **	6.11**	-1.86**	-0.31
Difference	(6.76)	(6.01)	(9.06)	(7.01)	(9.11)	(5.71)
Distance	-12.7%	-13.3%	-27.3%	37.2%	-6.8%	-3.3%
Difference						
(%)						

Table: Optimal Matching Distance

**p < 0.05, *p < 0.1

ELE SQC

→ < ∃ →</p>

Conclusions

- Hard to measure long time impact of score ranking rules
- Measure the bias of Yahoo's search engine towards advertisers and/or users at scale
- Analyze the relationship between users' and advertisers' preference
- Caveats:
 - Users' heterogeneity (Keyword shows clear intent)
 - Advertisers' targeting criteria (e.g. geo)
- We did not discuss:
 - Revenue implications
 - Simulate allocation mechanisms (e.g. VCG implementation Facebook)

(日本)

Thank You!

Migue	l Alc	ob	end	as

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □

Users' Utility

- Utility-maximizing model (allow counterfactuals)
- The model accommodates multiple clicks and unobserved correlation across ads within the impression.
- $u_{a(n)}(x)$: Perceived utility of ad *a* in position *n* with features x
- f_n : Cost of position n
- Infer $u_{a(n)}(x)$ and $f_n \forall n, a$
- Markov Decision Process (alternative logistic model):
 - State: displayed ads in an impression
 - Actions: clicks on ads or conclude search

$$V(\mathcal{C}_{t}, d_{t}) = \underbrace{\left(\sum_{n \in \mathcal{C}(t) \cup d_{t}} u_{a(n)}(x)^{1+R}\right)^{1/(1+R)} - \sum_{n \in \mathcal{C}(t) \cup d_{t}} f_{n} + \right.}_{Present Reward}_{\frac{\beta \log\left[\sum_{d_{t+1} \in D(t+1)} exp[V(\mathcal{C}_{t+1}, d_{t+1})]\right]}{Expected Future Reward}}$$

ELE NOR

Users' Utility: Estimation

- For a given value of u, f, α and R compute the continuation value V(.) for all t. In finite-horizon problems the optimal decision rule δ^{*} = (δ₀, ..., δ_T) is computed by backward induction starting at the terminal period T.
- 2 Given $V(\cdot)$ for all t and all d_t , compute the probability that users choose action d_t ,

$$P(d_t | C_t) = \frac{exp[V(C_t, d_t)]}{\sum_{l_t \in D(C_t)} exp[V(C_t, l_t)]}$$

 \bigcirc Iterate in order to find u, f, α and R that maximizes the likelihood function

$$\{\hat{u}, \hat{f}, \hat{\alpha}, \hat{R}\} = \operatorname{argmax}_{\theta} \prod_{k=1}^{K} \prod_{t=1}^{T_k} P(d_t^k | \mathcal{C}_t^k)$$

where K corresponds to the total number of impressions, and T_k is equal to the maximum number of sponsored ads in impression k.

Mi	one	lco	bend	las
	Buc	 ico	DCIN	445

▲□▶ ▲□▶ ▲□▶ ▲□▶ 三回 ののの

Results: Counterfactual

Table: Counterfactual

	Cheap Flight	Game	Мар
Counterfactual:			
Advertiser			
eCTR	25.32% **	45.21% **	-12% **
User Utility	18.16%*	20.67%	-1.53%
Advertiser Utility	26.35% **	77.29% **	4.42%
Counterfactual: User			
eCTR	43.81% **	73.34% **	6.39% *
User Utility	23.01%**	36.81%	3.26%
Advertiser Utility	13.56%	59.90% **	-8.67%

** p < 0.05, * p < 0.1

イロト イヨト イヨト イヨト

三日 のへで