

Real-Time Bidding based Display Advertising: Mechanisms and Algorithms

Shuai Yuan, MediaGamma Ltd

Weinan Zhang, UCL

Jun Wang, UCL

Shuai.yuan@mediagamma.com

{w.zhang,j.wang}@cs.ucl.ac.uk

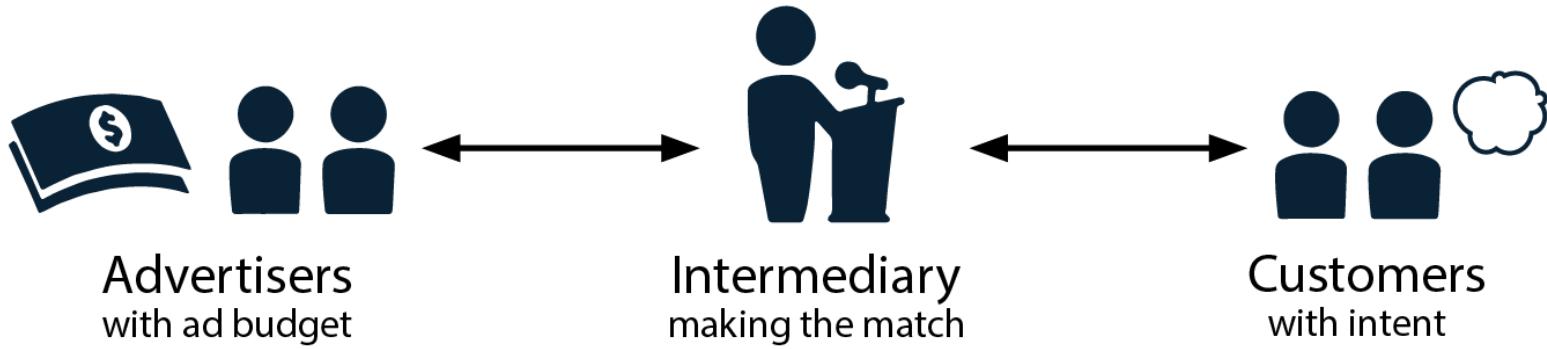
Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Table of contents

- RTB system
 - Auction mechanisms
 - User response estimation
 - Conversion attribution
 - Learning to bid
 - Data Management Platform (DMP) techniques
 - Floor price optimisation
 - Fighting against fraud

Advertising



- Make the best match between **advertisers** and **customers** with **economic constraints**



*“Half the money I spend
on advertising is wasted;
the trouble is I don’t
know which half.”*

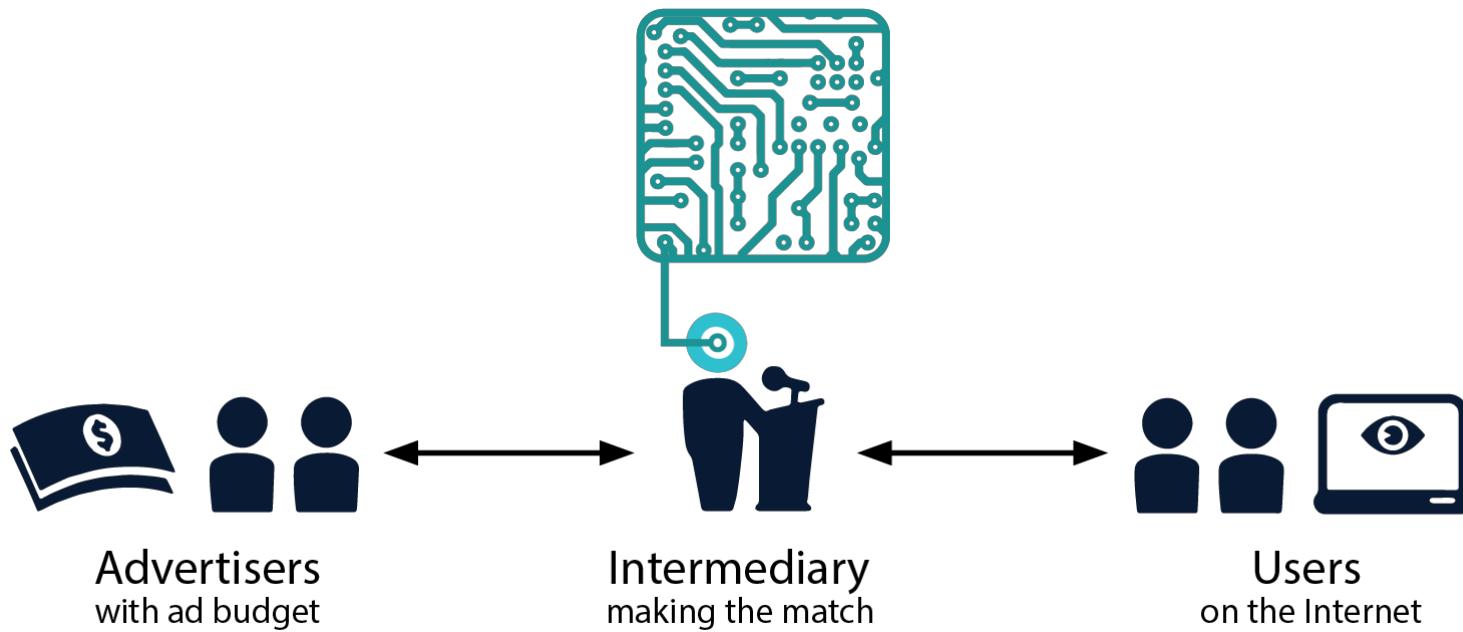
- John Wanamaker
(1838-1922)

*Father of modern advertising
and a pioneer in marketing*

Wasteful Traditional Advertising



Computational Advertising



- Design **algorithms** to make the best match between the advertisers and Internet users with economic constraints

Sponsored Search

Google | iphone 6s case

Web Shopping News Images Videos More ▾ Search tools

About 16,900,000 results (0.33 seconds)

[iPhone 6s Cases - case-mate.com](#)
Ad www.case-mate.com/iPhone-6s-Cases ▾
4.6 ★★★★★ rating for case-mate.com
Shop The iPhone 6s Case Collection. Free Standard Shipping!
Refined Protection · Slim & Tough · Genuinely Crafted · Premium Designs

[iPhone 6s](#)
Ad www.apple.com/ ▾
The only thing that's changed is everything. Learn more.
A9 chip · Two sizes · Now in rose gold
Pre-order 9.12 · iPhone Upgrade Program · 3D Touch · Cameras

In the news



[Speck's iPhone 6s CandyShell + MightyShell cases bring best-of-breed protection to Apple's latest iPhones](#)
9 to 5 Mac - 1 day ago
With the iPhone 6s and iPhone 6s Plus debuting next week, it's important to start thinking ...



[Shop for iphone 6s case on Google](#)

 Case-mate - Karat Case Fo... \$49.99 Best Buy ★★★★★ (163)	 Moshi - Iglaize Armour Case... \$39.99 Best Buy ★★★★★ (161)	 Logitech - Protection... \$21.99 Best Buy ★★★★★ (90)	 Moshi - Overture Wall... \$49.99 Best Buy ★★★★★ (18)
 Case-mate - Brilliance Cas... \$44.99 Best Buy ★★★★★ (294)	 Case-mate - Wallet Folio C... \$54.99 Best Buy ★★★★★ (173)	 Marc by Marc Jacobs Metalli... \$38.00 shopbop ★★★★★ (173)	 Case-mate - Karat Hard Sh... \$49.99 Best Buy ★★★★★ (34)

Search: iphone 6s case

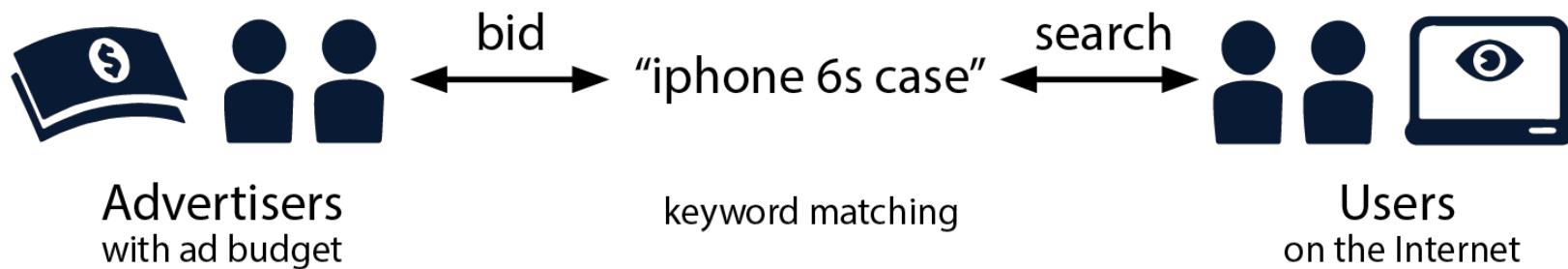
iPhone 6s Cases & Covers from OtterBox

www.otterbox.com/en-us/iphone-6s-cases ▾ OtterBox
Get protection that inspires confidence with iPhone 6s cases and covers from OtterBox.
Demandware SiteGenesis.

iPhone 6s - Accessories - Apple

www.apple.com › iPhone › iPhone 6s ▾ Apple Inc. ▾
The essential Apple-designed cases, accessories and all-new aluminum docks for iPhone 6s and iPhone 6s Plus.

Sponsored Search



- Advertiser sets a bid price for the keyword
- User searches the keyword
- Search engine hosts the auction to ranking the ads

Display Advertising

≡ The New York Times

SUBSCRIBE NOW

SIGN IN

Register

INTERNATIONAL | DEALBOOK | MARKETS | ECONOMY | ENERGY | MEDIA | TECHNOLOGY | PERSONAL TECH | ENTREPRENEURSHIP |

Exxon Mobil Investigated in New York Over Possible Lies on Climate

By JUSTIN GILLIS and CLIFFORD KRAUSS
3:30 PM ET

The sweeping inquiry, by the state attorney general, focuses on whether the oil company lied to the public and investors over the risks of climate change.

■ 250 Comments



T. Fallon/Bloomberg, via Getty Images

An Exxon Mobil refinery in Los Angeles, Calif. The New York attorney general is investigating the oil and gas company.

European Union Predicts Economic Gains From Influx of Migrants

By JAMES KANTER
12:10 PM ET

Officials forecast that the three million arrivals expected by 2017 would provide a net gain of perhaps a quarter of 1 percent by that year to the European economy.



INSIGHT & ANALYSIS

COMMON SENSE

Dewey Jury's Deadlock Exposes a System's Flaws

By JAMES B. STEWART
3:06 PM ET

One reason for the mistrial in the Dewey & LeBoeuf criminal case may have been the requirement for a unanimous decision.



LATEST NEWS

- | | |
|------------|---|
| 5:01 PM ET | 'Grand Theft Auto' Maker Take-Two's Revenue Nearly Triples |
| 5:00 PM ET | United Airlines CEO to Return in Early 2016 After Heart Attack |
| 4:57 PM ET | NY Attorney General Investigating Exxon Over Climate Statements |

MARKETS »

At close 11/05/2015

 BACKBASE

Backbase a Leader in the Forrester Wave for Omni-Channel Digital Banking

Read the Report



<http://www.nytimes.com/>

Internet Advertising Frontier: Real-Time Bidding (RTB) based Display Advertising

What is Real-Time Bidding?

- Every online **ad view** can be evaluated, bought, and sold, all **individually**, and all **instantaneously**.
- Instead of buying keywords or a bundle of ad views, advertisers are now **buying users** directly.

	DSP/Exchange	daily traffic
Advertising	iPinYou, China	18 billion impressions
	YOYI, China	5 billion impressions
	Fikisu, US	32 billion impressions
Finance	New York Stock Exchange	12 billion shares daily
	Shanghai Stock Exchange	14 billion shares daily

	Query per second
Turn DSP	1.6 million
Google	40,000 search

Suppose a student regularly reads articles on [emarketer.com](#)

eMarketer.

Research Topics Products Why eMarketer Customer Stories Articles

Advertisers Continue Rapid Adoption of Programmatic Buying

By 2017, advertisers will spend more than \$9 billion on RTB

Nov 26, 2013

Share Print Email

Advertisers are spending more than expected on real-time bidding, which is expected to account for a significant share of all display ad spending in the US billions, % change and % of total digital display ad spending

Year	RTB digital display ad spending (billions)	% change
2012	\$1.92	13.0%
2013	\$3.37	19.0%
2014	\$4.66	22.0%
2015	\$6.15	31.9%
2016	\$7.83	25.0%
2017	\$9.03	15.3%

eMarketer projects RTB digital display ad spending in the US will account for 29.0% of total US digital display ad spending by 2017, or \$9.03 billion. In 2013, it will account for 19.0%, or \$3.37 billion. These estimates are revised slightly upward from our previous forecast in August

Note: Includes all display formats served to all devices
Source: eMarketer, Dec 2013

Latest from eMarketer

Latest Articles Latest Webinars

Hispanic Gen Xers Lead in Daily Tablet Usage

Chrysler's Multichannel Approach to Online Video Gets Greater Recall

Android Rules UK Smartphone Sales

[More Articles »](#) [eMarketer Daily Newsletter »](#)

MARKETING PROGRAMS FOR EMAIL MARKETERS

FREE DOWNLOAD

WATCH THE VIDEO.

DO WHAT CAN NOW BE DONE. ☺

Content-related ads

He recently checked the London hotels

Booking.com

Browse by destination theme Shopping Fine Dining Culture Sightseeing Monuments Relaxation

home → uk → greater london → london → search results
16,378 properties 1,824 properties 1,574 properties London, 2 adults, 11 nights (Jul 14 - Jul 25) Change dates

(In fact, no login is required)

Search

Destination/Hotel Name:

Distance: 16 miles

Check-in Date: Mon 14 July 2014

Check-out Date: Fri 25 July 2014

I don't have specific dates yet

Guests: 2 Adults (1 room)

Search

Weinan Zhang 

London is a top choice with fellow travelers on your selected dates (48% reserved).
Tip: Prices might be higher than normal, so try searching with different dates if possible.

Try previous week Jul 7 - Jul 18 **Try next week** Jul 21 - Aug 1

930 out of 1857 properties are available in and around London
Showing 1 – 15

Sort by: Recommended Stars Location Price Review Score

 **Park Plaza Victoria London** ★★★★  1736
Central London, Westminster, London •  **Nearby stop**

Very good 8.5
Score from 1137 reviews

There are 13 people looking at this hotel.
Latest booking: 1 hour ago

Price for 11 nights £2,353.65
Superior Double Room We have 5 rooms left!
7 more room types  **Book now**

 **Central Park Hotel** ★★★★  1993

6.6

Relevant ads on facebook.com

Search for people, places and things

Weinan Home

Family
UCL
SJTU 16
UCL 20+
Shanghai Jiao Ton... 16
London, United Ki... 20+
University College... 20+
Close Friends
Intern,Beijing,Microso...

GROUPS
Microsoft Research C...
Create group

INTERESTS
Pages and Public Fig...

PAGES
Like Pages 1
Pages feed 9
Create a Page...

DEVELOPER

Secret Escapes Sponsored · *

Find the best rates on handpicked hotels



Secret Escapes | Exclusive Discounts
Get up to 70% off luxury hotels and holidays.

WWW.SECRETESCAPES.COM

Sign Up

Like · Comment · Share · 2,327 85 444

Bingkai Lin 43 mutual friends
Add Friend

Zhaomeng Peng 10 mutual friends
Add Friend

SPONSORED See all
247 London Hostel booking.com
Book & Save! 247 London Hostel, London.

Stale Marketing Stinks emarketer.com
 Freshen up with eMarketer's reports, trends & data on digital marketing. Download Today!

English (UK) · Privacy · Terms · Cookies · More ▾

Even on supervisor's homepage!

(User targeting dominates the context)

DR. JUN WANG
Computer Science, UCL

About Me Contact Publications Teaching Research Team Prospective Students Type text to search here...

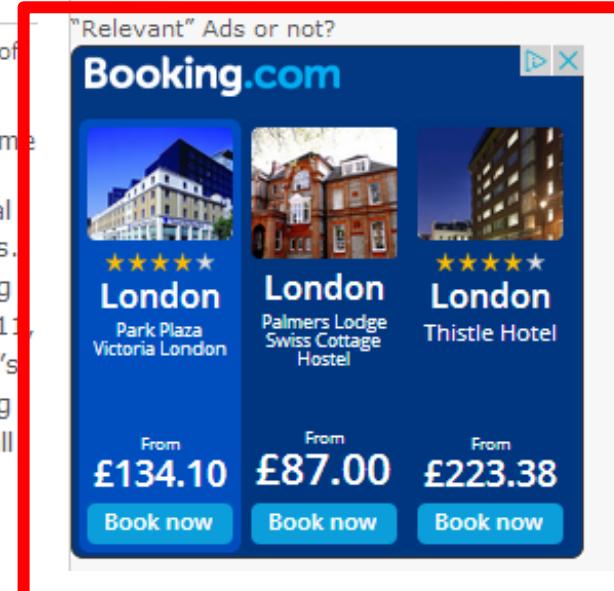
CIKM2013 Tutorial: Real-Time Bidding: A New Frontier of Computational Advertising Research

July 30th, 2013 Comments of

Online advertising is now one of the fastest advancing areas in IT industry. In display and mobile advertising, the most significant development in recent years is the growth of Real-Time Bidding (RTB), which allows selling and buying online display advertising in real-time one ad impression at a time. Since then, RTB has fundamentally changed the landscape of the digital media market by scaling the buying process across a large number of available inventories. It also encourages behaviour (re-)targeting, and makes a significant shift toward buying focused on user data, rather than contextual data. A report from IDC shows that in 2011, global RTB based display ad spend increased by 237% compared to 2010, with the U.S.'s \$2.2 billion RTB display spend leading the way. The market share of RTB-based spending of all display ad spending will grow from 10% in 2011 to 27% in 2016, and its share of all indirect spending will grow from 28% to 78%.

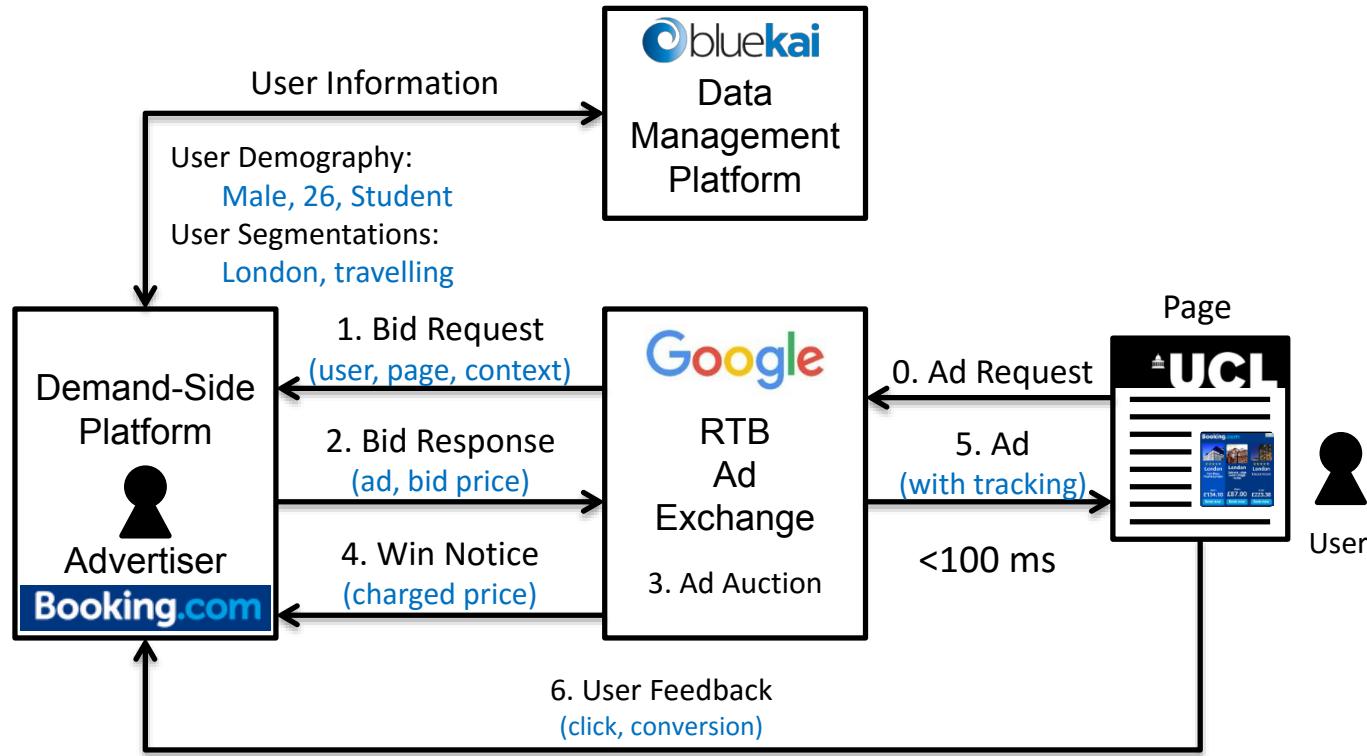
Scientifically, the further demand for automation, integration and optimization in RTB brings

"Relevant" Ads or not?



The image shows a Booking.com advertisement for travel deals to London. It features three hotel options: Park Plaza Victoria London (4 stars, From £134.10), Palmers Lodge Swiss Cottage Hostel (4 stars, From £87.00), and Thistle Hotel (4 stars, From £223.38). Each listing includes a 'Book now' button.

RTB Display Advertising Mechanism



- Buying ads via real-time bidding (RTB), 10B per day

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Auctions scheme

private values bids



v_1

$\rightarrow b_1$



v_2

$\rightarrow b_2$



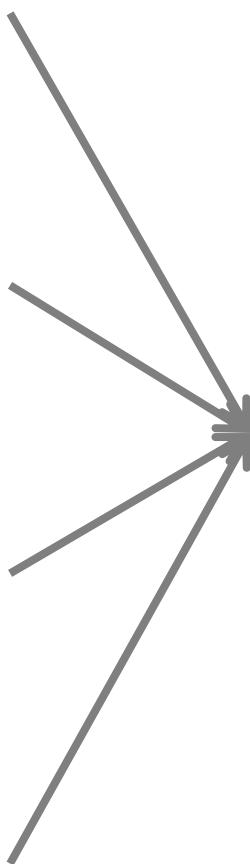
v_3

$\rightarrow b_3$



v_4

$\rightarrow b_4$



winner



payments \$\$\$

Modeling



- n bidders
- Each bidder has value v_i for the item
 - “willingness to pay”
 - Known only to him – “private value”
- If bidder i wins and pays p_i , his utility is $v_i - p_i$
 - In addition, the utility is 0 when the bidder loses.
- Note: bidders prefer losing than paying more than their value.

Strategy

- A strategy for each bidder
 - how to bid given your intrinsic, private value?
 - a strategy here is a *function*, a plan for the game.
Not just a bid.
- Examples for strategies:
 - $b_i(v_i) = v_i$ (truthful)
 - $b_i(v_i) = v_i/2$
 - $b_i(v_i) = v_i/n$
 - If $v < 50$, $b_i(v_i) = v_i$
otherwise, $b_i(v_i) = v_i + 17$
- Can be modeled as normal form game, where these strategies are the pure strategies.
- Example for a *game with incomplete information*.

	$B(v)=v$	$B(v)=v/2$	$B(v)=v/n$
$B(v)=v$				
...				

Strategies and equilibrium

- An equilibrium in the auction is a profile of strategies B_1, B_2, \dots, B_n such that:
 - Dominant strategy equilibrium: each strategy is optimal whatever the other strategies are.
 - Nash equilibrium: each strategy is a best response to the other strategies.

	$B(v)=v$	$B(v)=v/2$	$B(v)=v/n$
$B(v)=v$				
...				

Bayes-Nash equilibrium

- Recall a set of bidding strategies is a **Nash equilibrium** if each bidder's strategy maximizes his payoff given the optimal strategies of the others.
 - In auctions: bidders do not know their opponent's values, i.e., there is *incomplete information*.
 - Each bidder's strategy must maximize her *expected* payoff accounting for the uncertainty about opponent values.

1st price auctions

- $\text{Truthful}(b_i = v_i) ? \text{ NO!}$



Equilibrium in 2rd-price auctions

- bidder 1's payoff

$$\begin{cases} v_1 - b_i & \text{if } b_1 > b_i > \max\{b(v_2), \dots, b(v_{i-1}), b(v_{i+1}), \dots, b(v_n)\} \\ 0 & \text{if } b_1 \leq \max\{b(v_2), \dots, b(v_n)\} \end{cases}$$

- The expected payoff of bidding b_1 is given by

$$\pi(v_1, b_1) = \int_0^{b_1} (v_1 - x) dF^{N-1}(x) = \int_0^{b_1} (N-1)(v_1 - x) f(x) F^{N-2}(x) dx$$

- Suppose $b_1 < v_1$, if b_1 is increased to v_1 the integral increases by the amount

$$\int_{b_1}^{v_1} (N-1)(v_1 - x) f(x) F^{N-2}(x) dx$$

- The reverse happens if $b_1 > v_1$

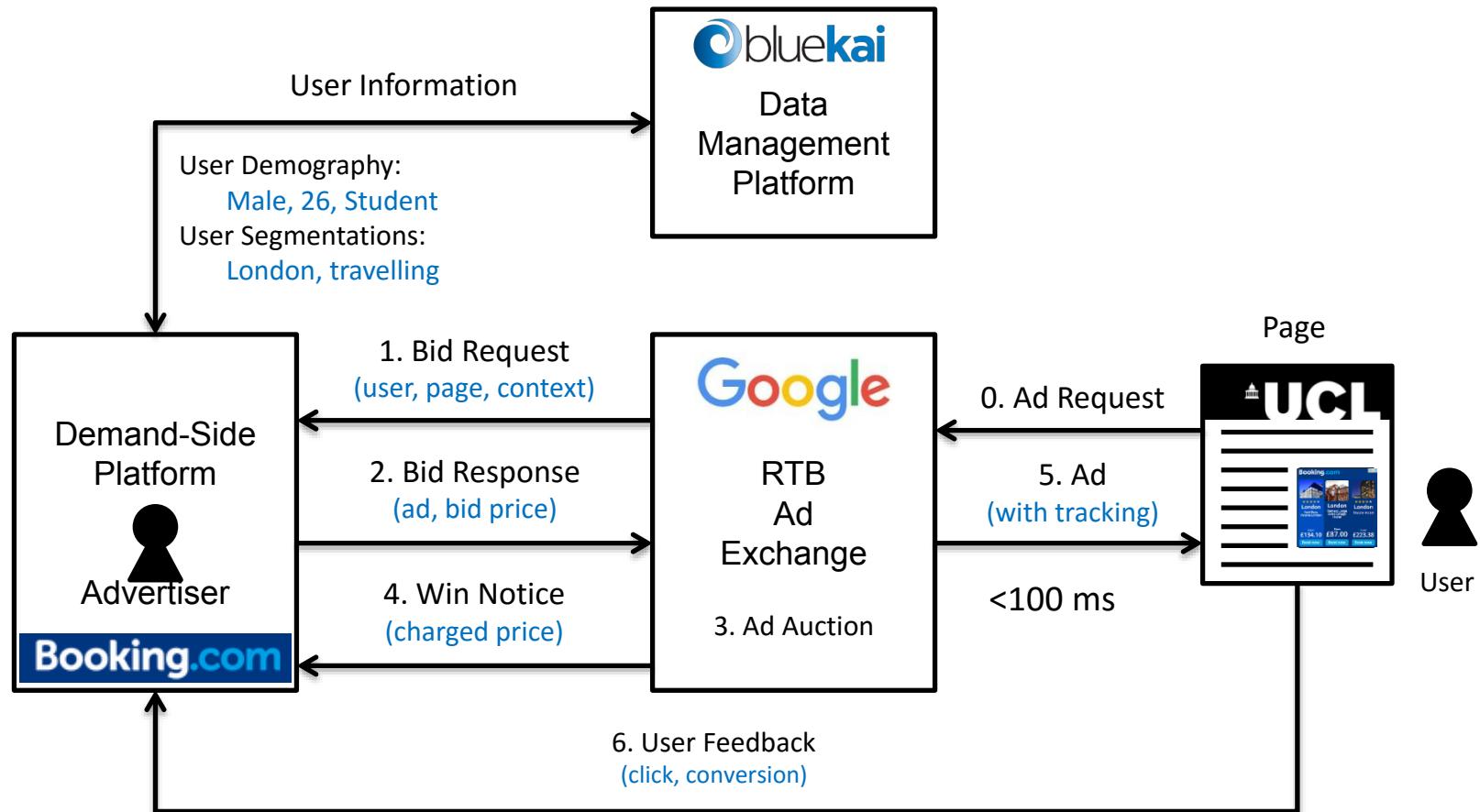
Reserve Prices and Entry Fees

- *Reserve Prices*: the seller is assumed to have committed to not selling below the reserve
 - Reserve prices are assumed to be known to all bidders
 - The reserve prices = the minimum bids
- *Entry Fees*: those bidders who enter have to pay the entry fee to the seller
- They reduce bidders' incentives to participate, but they might increase revenue as 1) the seller collects extra revenues 2) bidders might bid more aggressively

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

RTB Display Advertising Mechanism



- Buying ads via real-time bidding (RTB), 10B per day

Predict how likely the user is going to click the displayed ad.

≡ Q The New York Times

SUBSCRIBE NOW

SIGNIN

Register



INTERNATIONAL | DEALBOOK | MARKETS | ECONOMY | ENERGY | MEDIA | TECHNOLOGY | PERSONAL TECH | ENTREPRENEURSHIP |

Exxon Mobil Investigated in New York Over Possible Lies on Climate

By JUSTIN GILLIS and CLIFFORD KRAUSS
3:30 PM ET

The sweeping inquiry, by the state attorney general, focuses on whether the oil company lied to the public and investors over the risks of climate change.

■ 250 Comments



T. Fallon/Bloomberg, via Getty Images

An Exxon Mobil refinery in Los Angeles, Calif. The New York attorney general is investigating the oil and gas company.

European Union Predicts Economic Gains From Influx of Migrants

By JAMES KANTER
12:10 PM ET

Officials forecast that the three million arrivals expected by 2017 would provide a net gain of perhaps a quarter of 1 percent by that year to the European economy.



INSIGHT & ANALYSIS

COMMON SENSE

Dewey Jury's Deadlock Exposes a System's Flaws

By JAMES B. STEWART
3:06 PM ET

One reason for the mistrial in the Dewey & LeBoeuf criminal case may have been the requirement for a unanimous decision.



LATEST NEWS

- 5:01 PM ET 'Grand Theft Auto' Maker Take-Two's Revenue Nearly Triples
- 5:00 PM ET United Airlines CEO to Return in Early 2016 After Heart Attack
- 4:57 PM ET NY Attorney General Investigating Exxon Over Climate Statements

MARKETS »

At close 11/05/2015

 BACKBASE

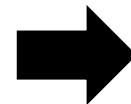
Backbase a Leader in the Forrester Wave for Omni-Channel Digital Banking



User response estimation problem

- Click-through rate estimation as an example

- Date: 20160320
- Hour: 14
- Weekday: 7
- IP: 119.163.222.*
- Region: England
- City: London
- Country: UK
- Ad Exchange: Google
- Domain: yahoo.co.uk
- URL: <http://www.yahoo.co.uk/abc/xyz.html>
- OS: Windows
- Browser: Chrome
- Ad size: 300*250
- Ad ID: a1890
- User tags: Sports, Electronics



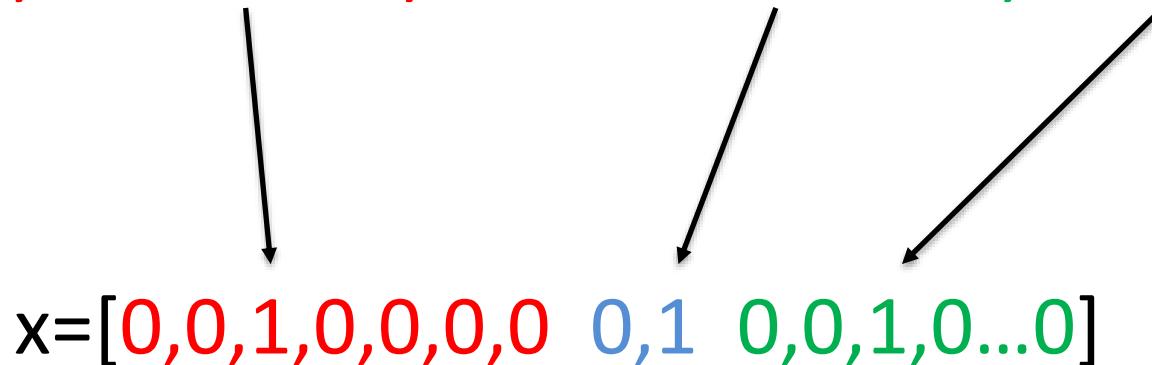
Click (1) or not (0)?

Predicted CTR (0.15)

Feature Representation

- Binary one-hot encoding of categorical data

$x = [\text{Weekday=Wednesday}, \text{Gender=Male}, \text{City=London}]$



High dimensional sparse binary feature vector

Linear Models

- Logistic Regression
 - With SGD learning
 - Sparse solution
- Online Bayesian Profit Regression

ML Framework of CTR Estimation

- A binary regression problem

$$\min_{\mathbf{w}} \sum_{(y, \mathbf{x}) \in D} \mathcal{L}(y, \hat{y}) + \lambda \Phi(\mathbf{w})$$

- Large binary feature space (>10 millions)
 - Bloom filter to detect and add new features (e.g., > 5 instances)
- Large data instance number (>10 millions daily)
- A seriously unbalanced label
 - Normally, #click/#non-click = 0.3%
 - Negative down sampling
 - Calibration

Logistic Regression

- Prediction

$$\hat{y} = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}}$$

- Cross Entropy Loss

$$\mathcal{L}(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$$

- Stochastic Gradient Descent Learning

$$\mathbf{w} \leftarrow (1 - \lambda)\mathbf{w} + \eta(y - \hat{y})\mathbf{x}$$

Logistic Regression with SGD

$$\mathbf{w} \leftarrow (1 - \lambda)\mathbf{w} + \eta(y - \hat{y})\mathbf{x}$$

- Pros
 - Standardised, easily understood and implemented
 - Easy to be parallelised
- Cons
 - Learning rate η initialisation
 - Uniform learning rate against different binary features

Logistic Regression with FTRL

- In practice, we need a sparse solution as >10 million feature dimensions
- Follow-The-Regularised-Leader (FTRL) online Learning

$$\mathbf{w}_{t+1} = \arg \min_{\mathbf{w}} \left(\mathbf{g}_{1:t} \cdot \mathbf{w} + \frac{1}{2} \sum_{s=1}^t \sigma_s \|\mathbf{w} - \mathbf{w}_s\|_2^2 + \lambda_1 \|\mathbf{w}\|_1 \right)$$

s.t. $\mathbf{g}_{1:t} = \sum_{s=1}^t \mathbf{g}_s$

$\sigma_s = \sqrt{s} - \sqrt{s-1}$

adaptively selects regularization functions

t: current example index
 \mathbf{g}_s : gradient for example t

- Online closed-form update of FTRL

$$w_{t+1,i} = \begin{cases} 0 & \text{if } |z_{t,i}| \leq \lambda_1 \\ -\eta_t(z_{t,i} - \text{sgn}(z_{t,i})\lambda_1) & \text{otherwise.} \end{cases}$$

$$\mathbf{z}_{t-1} = \mathbf{g}_{1:t-1} - \sum_{s=1}^{t-1} \sigma_s \mathbf{w}_s$$

$$\eta_{t,i} = \frac{\alpha}{\beta + \sqrt{\sum_{s=1}^t g_{s,i}^2}}$$

[McMahan et al. Ad Click Prediction : a View from the Trenches. KDD 13]

[Xiao, Lin. "Dual averaging method for regularized stochastic learning and online optimization." Advances in Neural Information Processing Systems. 2009]

Online Bayesian Probit Regression

Given feature x , predicting click y

$$p(y|x, \mathbf{w}) := \Phi\left(\frac{y \cdot \mathbf{w}^T \mathbf{x}}{\beta}\right)$$

Where probit function $\Phi(t) := \int_{-\infty}^t \mathcal{N}(s; 0, 1) ds$

And prior distribution $p(\mathbf{w}) = \prod_{i=1}^N \prod_{j=1}^{M_i} \mathcal{N}(w_{i,j}; \mu_{i,j}, \sigma_{i,j}^2)$

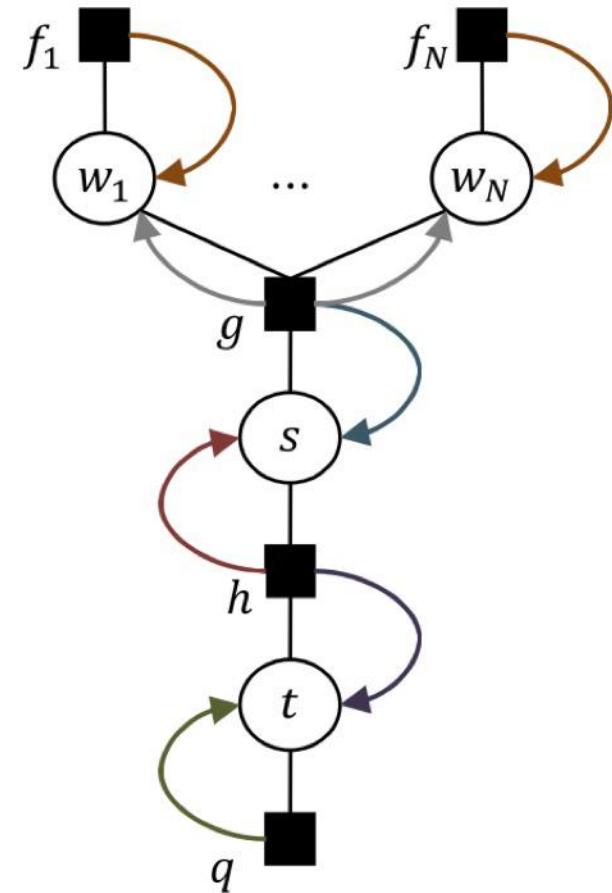
The factorised model

$$p(y | t) \cdot p(t | s) \cdot p(s | \mathbf{x}, \mathbf{w}) \cdot p(\mathbf{w})$$

Where $p(s|\mathbf{x}, \mathbf{w}) := \delta(s = \mathbf{w}^T \mathbf{x})$.

$$p(t|s) := \mathcal{N}(t; s, \beta^2)$$

$$p(y|t) := \delta(y = \text{sign}(t)).$$



Approximated inference via
Expectation Propagation

Linear Prediction Models

$$\hat{y} = f(\mathbf{w}^T \mathbf{x})$$

- Pros
 - Highly efficient and scalable
 - Explore larger feature space and training data
- Cons
 - Modelling limit: feature independence assumption
 - Cannot capture feature interactions unless defining high order combination features
 - E.g., hour=10AM & city=London & browser=Chrome

Non-linear Models

- Gradient Boosting Decision Trees
- Factorisation Machines
- Combined Models
- Deep Neural Networks

Factorisation Machines

- Prediction based on feature embedding

$$\hat{y}(\mathbf{x}) = \sigma \left(w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n x_i x_j \mathbf{v}_i^T \mathbf{v}_j \right)$$

- Explicitly model feature interactions
 - Second order, third order etc.
- Empirically better than logistic regression
- A new way for **user profiling**

[Rendle. Factorization machines. ICDM 2010.]

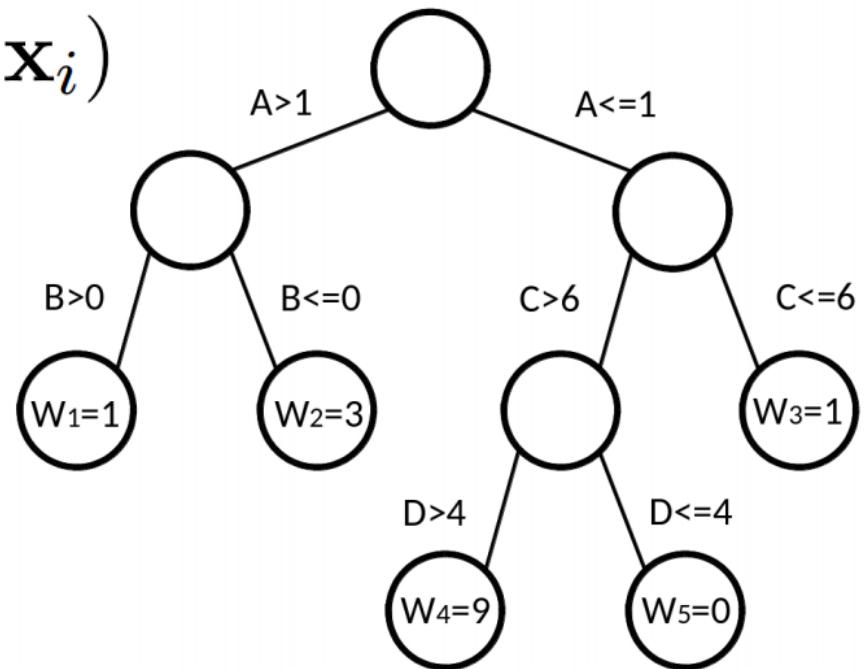
[Oentaryo et al. Predicting response in mobile advertising with hierarchical importance-aware factorization machine. WSDM 14]

Gradient Boosting Decision Trees

- Additive decision trees for prediction

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F}$$

- Each decision tree $f_k(\mathbf{x}_i)$



Gradient Boosting Decision Trees

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^K f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F}$$

- Learning

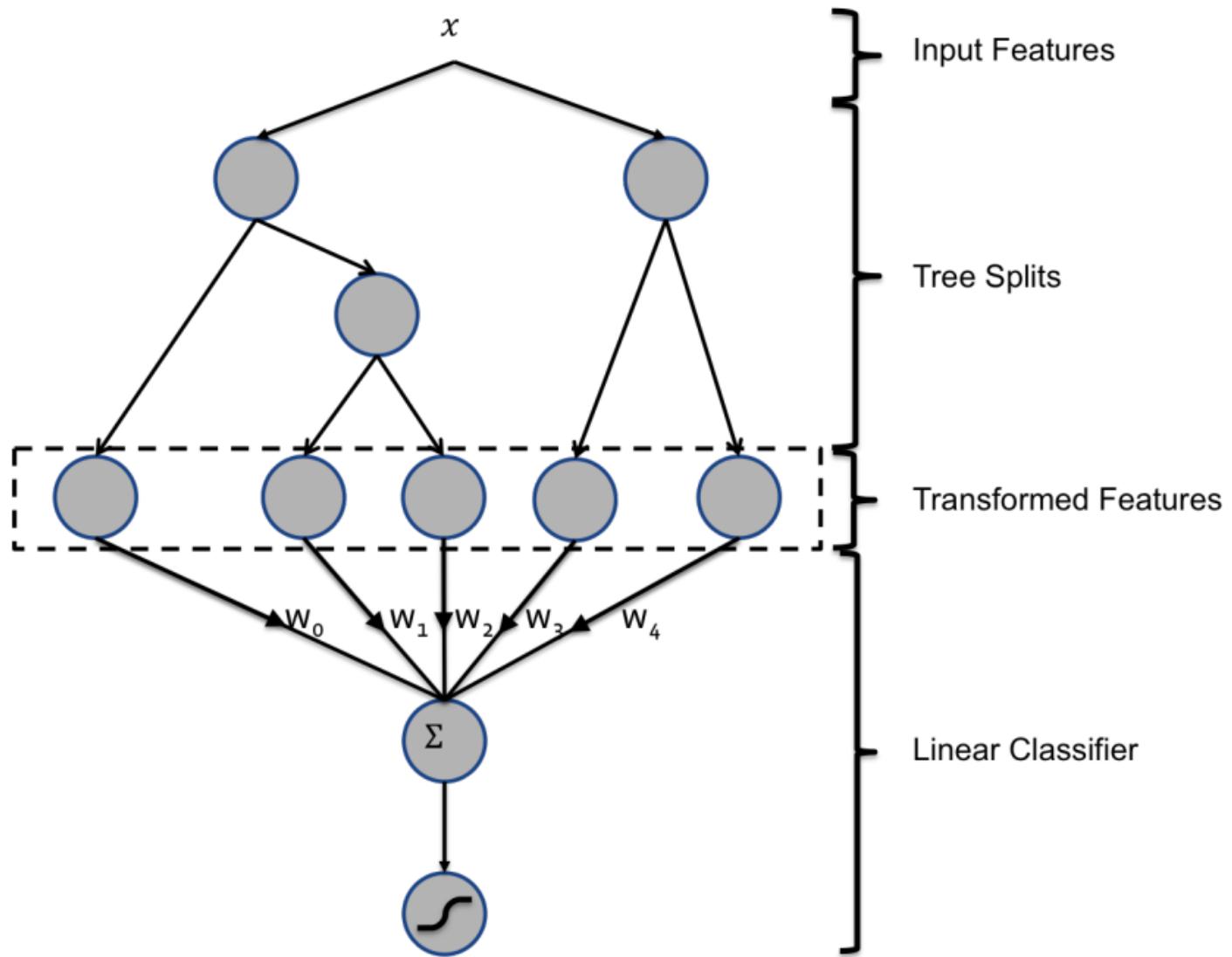
$$\mathcal{L}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_i)$$

$$= \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \sum_{i=1}^t \Omega(f_i)$$

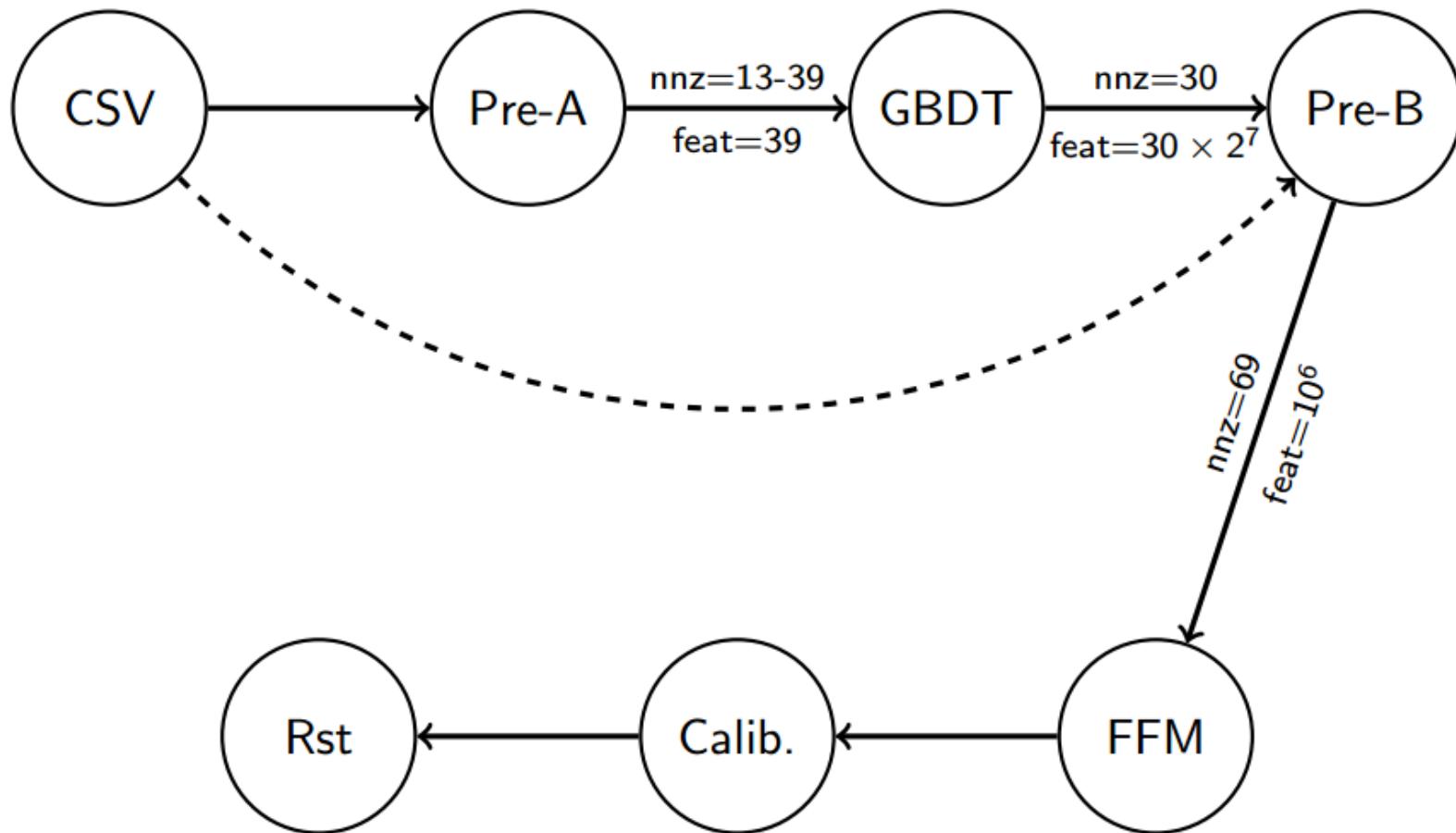
$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^n [l(y_i, \hat{y}^{(t-1)}) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i)] + \sum_{i=1}^t \Omega(f_i)$$

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \quad h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)})$$

Combined Models: GBDT + LR



Combined Models: GBDT + FM



"nnz" means the number of non-zero elements of each impression; "feat" represents the size of feature space.

CTR

Fully Connected

Hiden Layer (l_2)

Fully Connected

Hiden Layer (l_1)

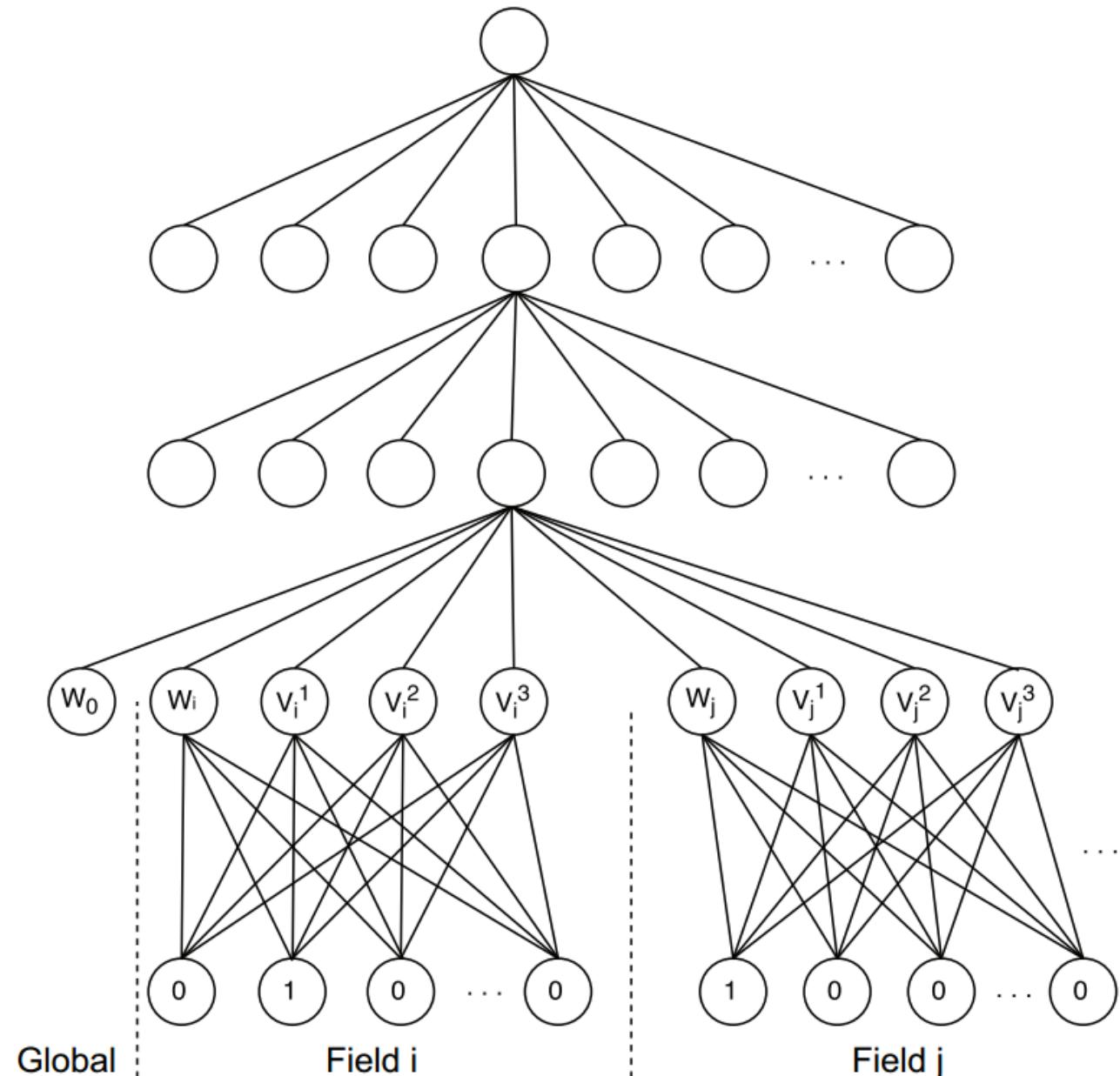
Fully Connected

Dense Real Layer (z)

Initialised by FM's
Weights and Vectors.

Fully Connected within
each field

Sparse Binary
Features (x)

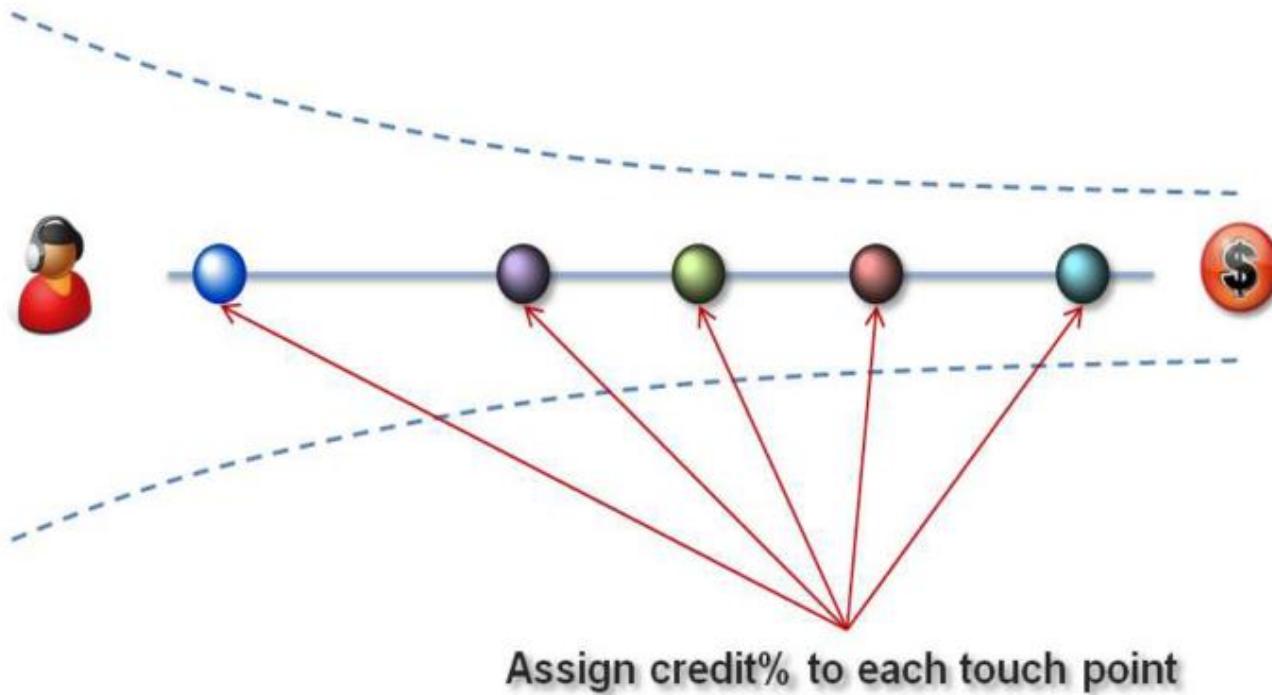


[Zhang et al. Deep Learning over Multi-field Categorical Data – A Case Study on User Response Prediction. ECIR 16] in Monday Machine Learning Track

Table of contents

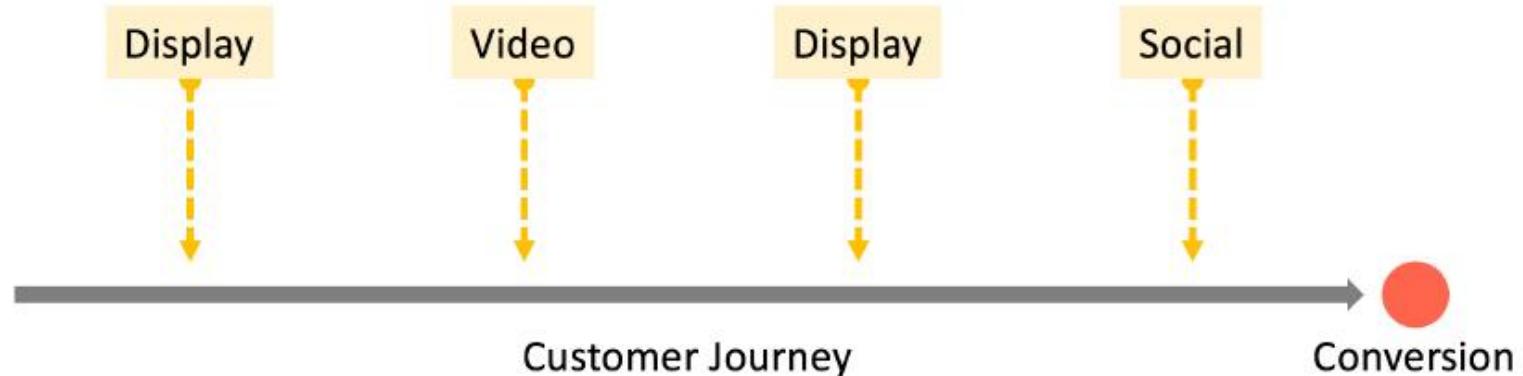
- RTB System
- Auction Mechanisms
- CTR Estimation
- **Conversion Attribution**
- Learning to Bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Conversion Attribution



- Assign credit% to each channel according to contribution
 - Current industrial solution: last-touch attribution
- [Shao et al. Data-driven multi-touch attribution models. KDD 11]

Heuristics-based Attribution



Model	Attribution			
Last Touch	0%	0%	0%	100%
First Touch	100%	0%	0%	0%
Linear	25%	25%	25%	25%
Time Decay	10%	20%	30%	40%
Position Based	40%	10%	10%	40%

[Kee. Attribution playbook – google analytics. Online access.]

A Good Attribution Model

- Fairness
 - Reward an individual channel in accordance with its ability to affect the likelihood of conversion
- Data driven
 - Using ad touch and conversion data for each campaign to build its model
- Interpretability
 - Generally accepted by all parties

Bagged Logistic Regression

Display	Search	Mobile	Email	Social	Convert?
1	1	0	0	1	1
1	0	1	1	1	0
0	1	0	1	0	1
0	0	1	1	1	0

- For M iterations
 - Sample 50% data instances and 50% features
 - Train a logistic regression and record the weights
- Average the feature weights

Bagged Logistic Regression

Display	Search	Mobile	Email	Social	Convert?
1	1	0	0	1	1
1	0	1	1	1	0
0	1	0	1	0	1
0	0	1	1	1	0

- For M iterations
 - Sample 50% data instances and 50% features
 - Train a logistic regression and record the weights
- Average the feature weights

Bagged Logistic Regression

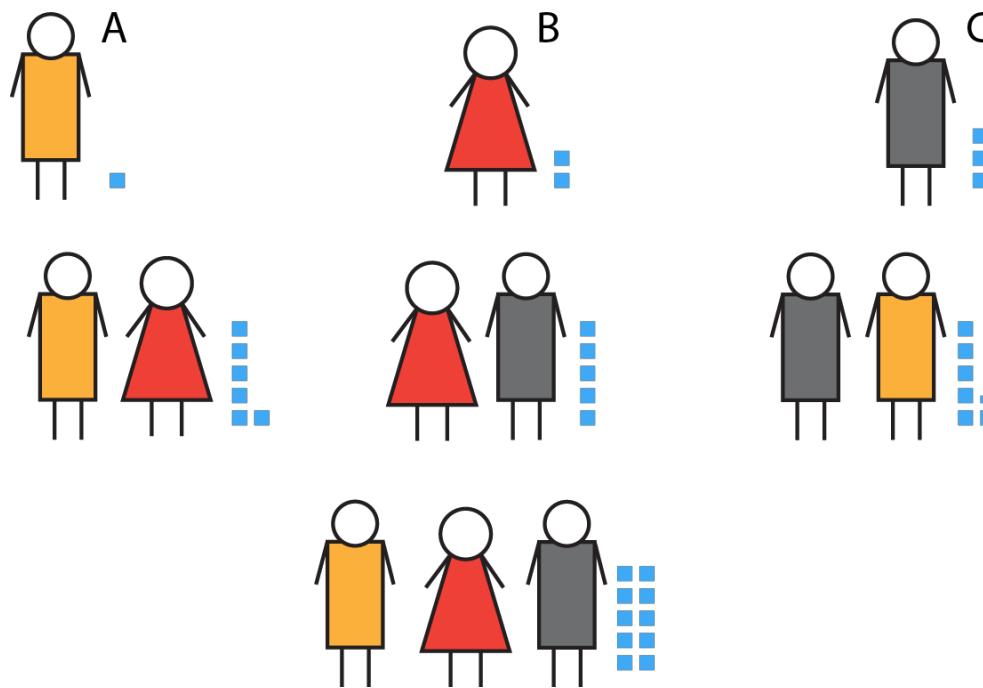
Display	Search	Mobile	Email	Social	Convert?
1	1	0	0	1	1
1	0	1	1	1	0
0	1	0	1	0	1
0	0	1	1	1	0

- For M iterations
 - Sample 50% data instances and 50% features
 - Train a logistic regression and record the weights
- Average the feature weights

[Shao et al. Data-driven multi-touch attribution models. KDD 11]

Shapley Value based Attribution

- Coalition game
 - How much does a player contribute in the game



[Fig source: <https://pjdelta.wordpress.com/2014/08/10/group-project-how-much-did-i-contribute/>]

Shapley Value based Attribution

- Coalition game

$$V_k = \sum_{S \subseteq C/k} \omega_{S,k} \cdot [E[\gamma|S \cup C_k] - E[\gamma|S]]$$

$$\omega_{S,k} = \frac{|S|!(|C| - |S| - 1)!}{|C|!}$$

A Probabilistic Attribution Model

- Conditional probabilities

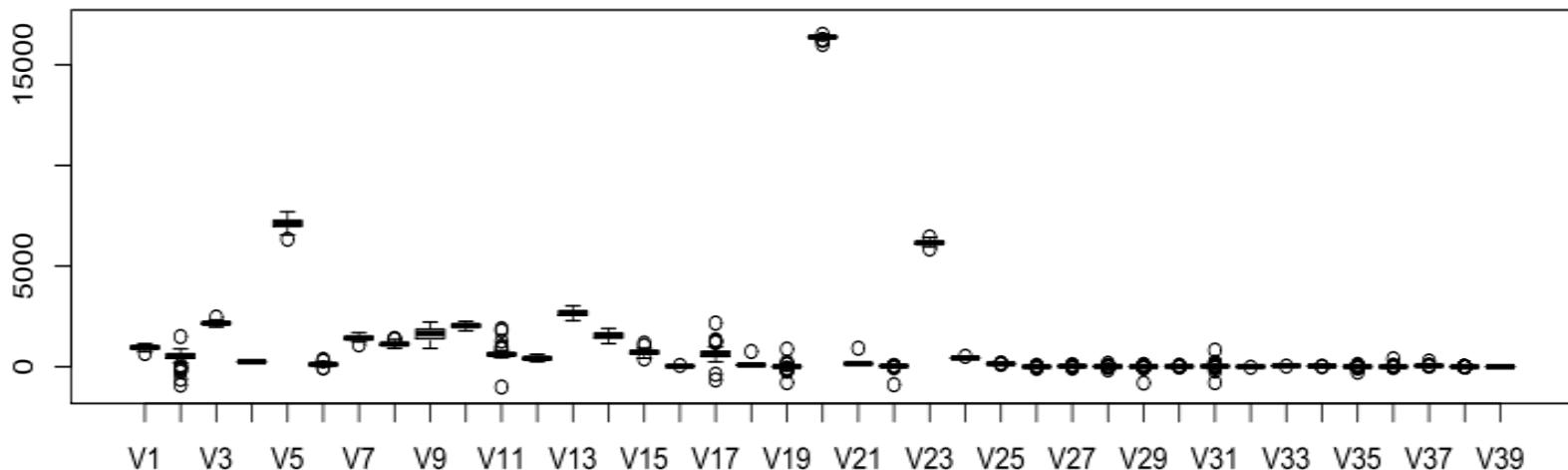
$$P(y|x_i) = \frac{N_{positive}(x_i)}{N_{positive}(x_i) + N_{negative}(x_i)}$$

$$P(y|x_i, x_j) = \frac{N_{positive}(x_i, x_j)}{N_{positive}(x_i, x_j) + N_{negative}(x_i, x_j)}$$

- Attributed contribution

$$V(x_i) = \frac{1}{2}P(y|x_i) + \frac{1}{2N_{j \neq i}} \sum_{j \neq i} \left(P(y|x_i, x_j) - P(y|x_j) \right)$$

bagged logistic regression model



simple probabilistic model

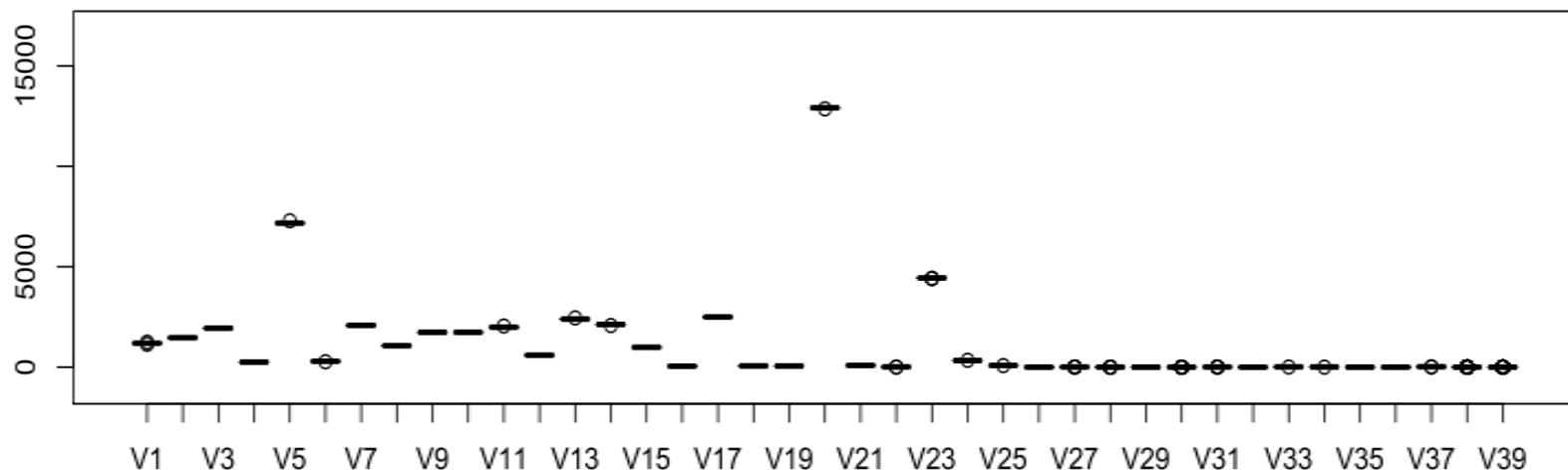


Table 2: The MTA user-level attribution analysis.

Channel	MTA Total	LTA Total	Difference
Search Click	17,494	17,017	97%
Email Click	6,938	7,340	106%
Display Network A	5,567	8,148	146%
Display Network G	2,037	470	23%
Display Network B	1,818	1,272	70%
Display Trading Desk	1,565	1,367	87%
Display Network C	1,494	1,373	92%
Display Network D	1,491	1,233	83%
Email View	1,420	458	32%
Display Network E	1,187	1,138	96%
Brand Campaign	907	1,581	174%
Social	768	1,123	146%
Display Network H	746	284	38%
Display Network F	673	787	117%
Display Network I	489	136	28%
Retail Email Click	483	491	102%
Display Network J	222	92	41%
Retail Email	168	110	66%
Social Click	133	153	115%
Video	58	31	54%

Data-Driven Probabilistic Models

- The “relatively heuristic” data-driven model
[Shao et al. Data-driven multi-touch attribution models. KDD 11]

$$V(x_i) = \frac{1}{2}P(y|x_i) + \frac{1}{2N_{j \neq i}} \sum_{j \neq i} \left(P(y|x_i, x_j) - P(y|x_j) \right)$$

- A more generalized and data-driven model
[Dalessandro et al. Casually Motivated Attribution for Online Advertising. ADKDD 11]

$$V(x_i) = \sum_{S \subseteq I \setminus i} w_{S,i} (P(y|S, x_i) - P(y|S))$$

- $w_{S,i}$: the probability that the sequence begin with (S, C_i)

Attribution Comparison

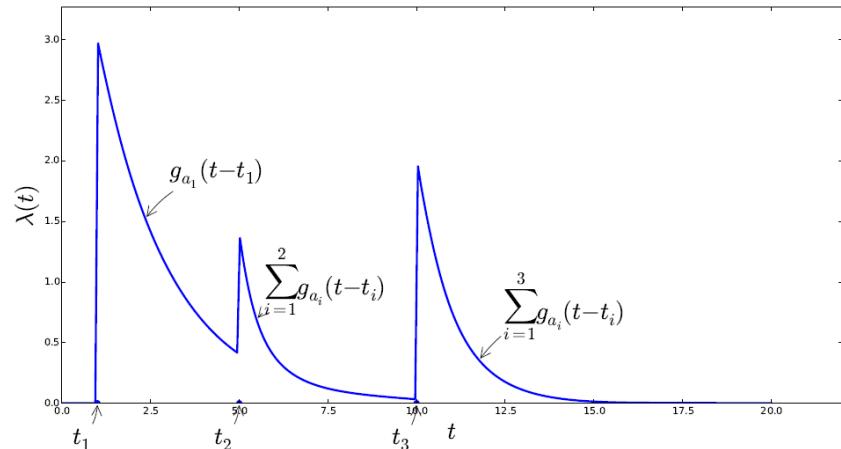
		Data Generating Parameters			Attribution Results			
Channel	Group	Ad Propensity Likelihood	Simulated Conversion Rate	Last Touch Propensity	Last Touch Conversions	Multi Touch Conversions	Delta N	Delta %
1	Gen Prospecting	5.0%	0.100%	0.2%	1,023	2,176	1,153	113%
2	Gen Prospecting	10.0%	0.080%	0.2%	1,932	3,284	1,352	70%
3	Gen Prospecting	10.0%	0.070%	0.2%	1,854	3,085	1,231	66%
4	Gen Prospecting	15.0%	0.050%	0.2%	2,491	3,434	943	38%
5	Gen Prospecting	15.0%	0.050%	1.8%	3,134	3,143	9	0%
6	Gen Prospecting	20.0%	0.010%	1.7%	2,998	736	-2,262	-75%
7	Gen Prospecting	20.0%	0.008%	6.7%	3,558	260	-3,298	-93%
8	Gen Prospecting	25.0%	0.008%	6.8%	4,406	409	-3,997	-91%
9	Retargeting	2.5%	0.500%	3.0%	3,921	5,673	1,752	45%
10	Retargeting	2.5%	0.400%	6.0%	3,375	4,489	1,114	33%
11	Retargeting	3.0%	0.300%	10.5%	3,468	4,068	600	17%
12	Retargeting	3.5%	0.250%	15.3%	3,728	3,997	269	7%
13	Search	0.5%	1.000%	23.7%	2,109	2,430	321	15%
14	Search	0.5%	2.000%	23.6%	5,329	5,045	-284	-5%

- Help find some “cookie bombing” channels

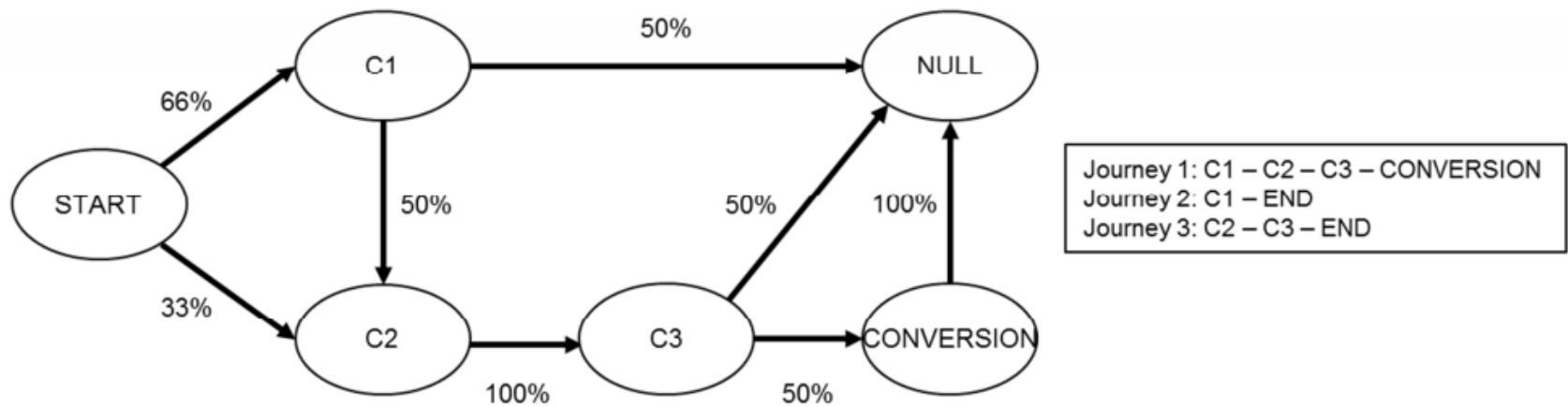
Other Attribution Models

- Survival models with time

[Zhang et al. Multi-Touch Attribution in Online Advertising with Survival Theory. ICDM 2014]



- Markov graph

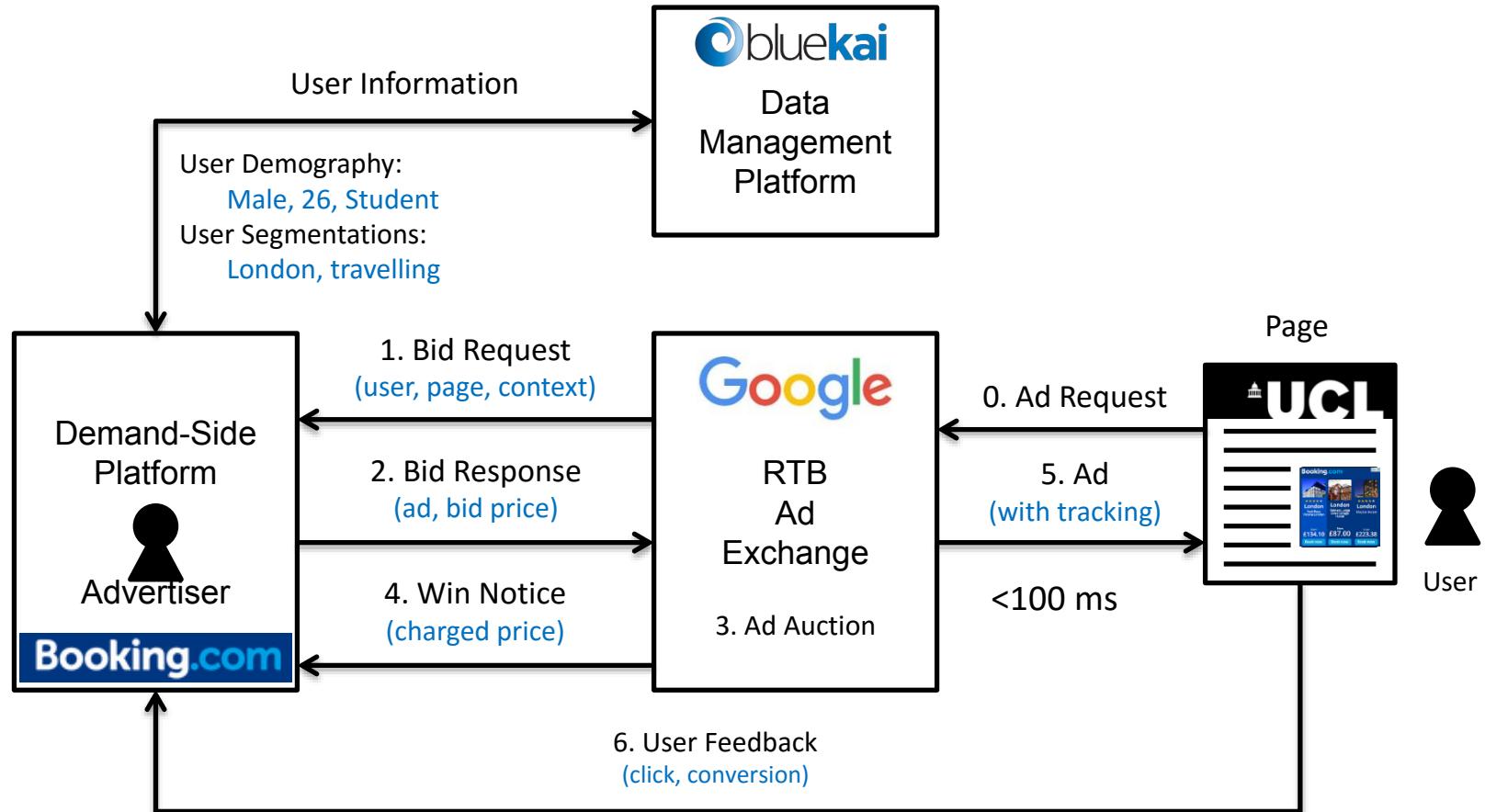


[Anderl et al. Mapping the customer journey: A graph-based framework for online attribution modeling. SSRN 2014]

Table of contents

- RTB System
- Auction Mechanisms
- CTR Estimation
- Conversion Attribution
- **Learning to Bid**
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

RTB Display Advertising Mechanism



- Buying ads via real-time bidding (RTB), 10B per day

Data of Learning to Bid

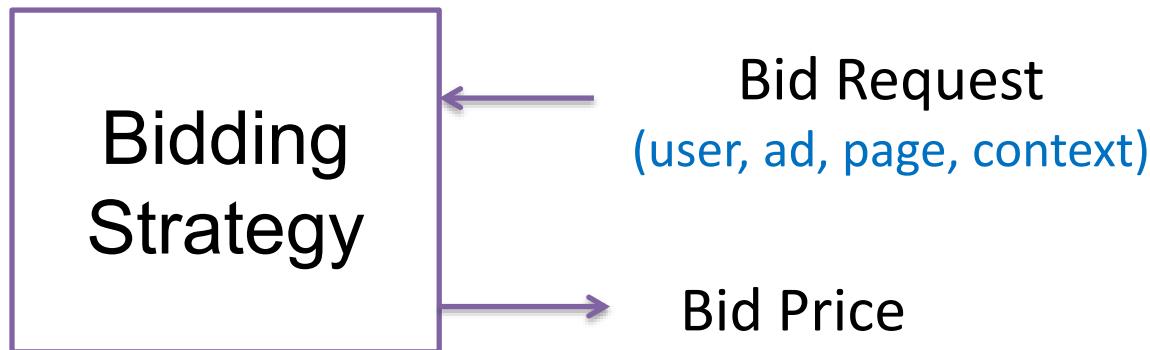
- Data

(\mathbf{x}, t)	b	w	c	y
(up, 1500×20, Shanghai, 0)	5	1	4	1
(down, 1200×25, Paris, 1)	4	1	3	0
(left, 20×1000, Los Angeles, 2)	3	0	×	×
(right, 35×600, London, 3)	0	0	×	×

- Bid request features: High dimensional sparse binary vector
- Bid: Non-negative real or integer value
- Win: Boolean
- Cost: Non-negative real or integer value
- Feedback: Binary

Problem Definition of Learning to Bid

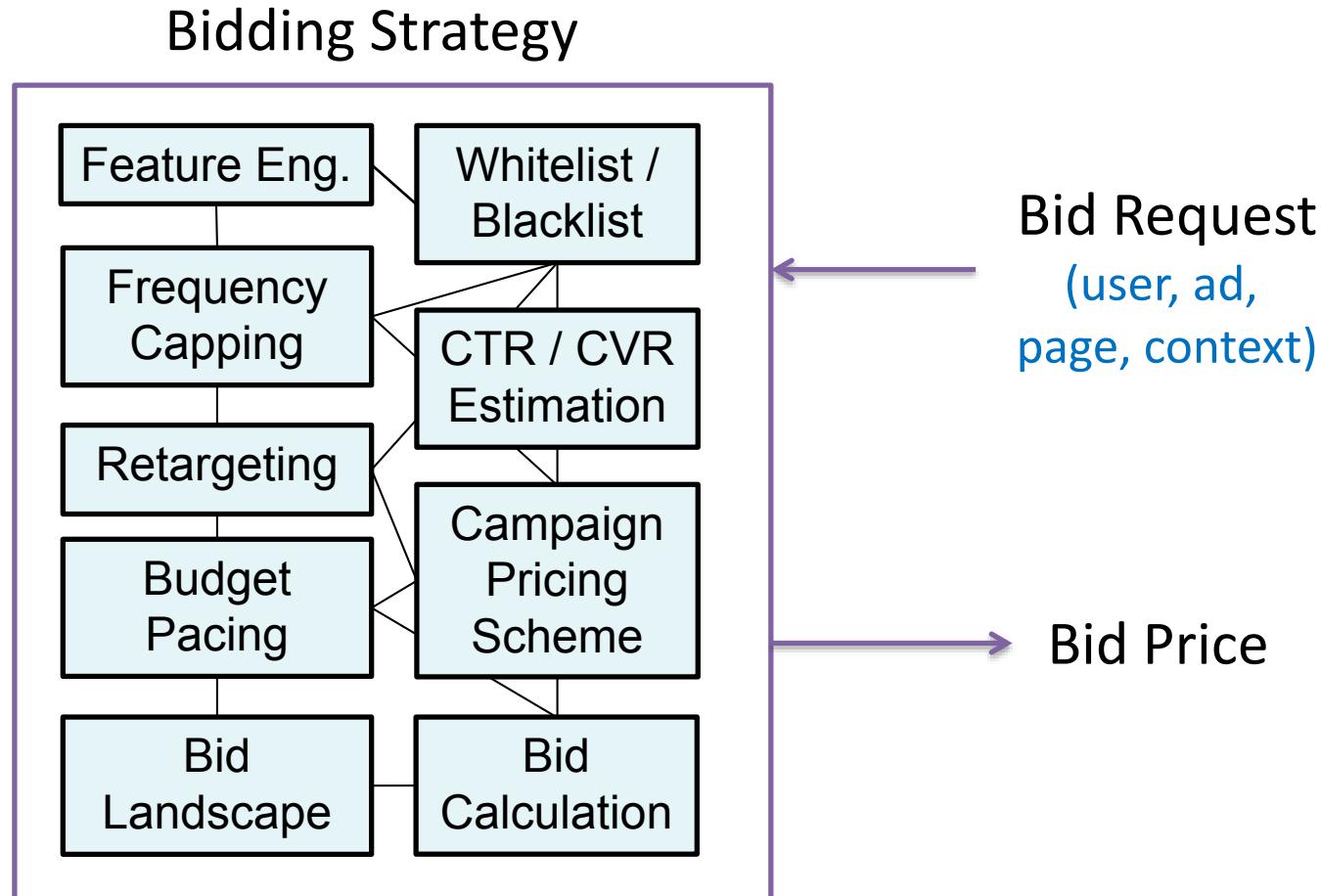
- How much to bid for each bid request?
 - Find an optimal bidding function $b(x)$



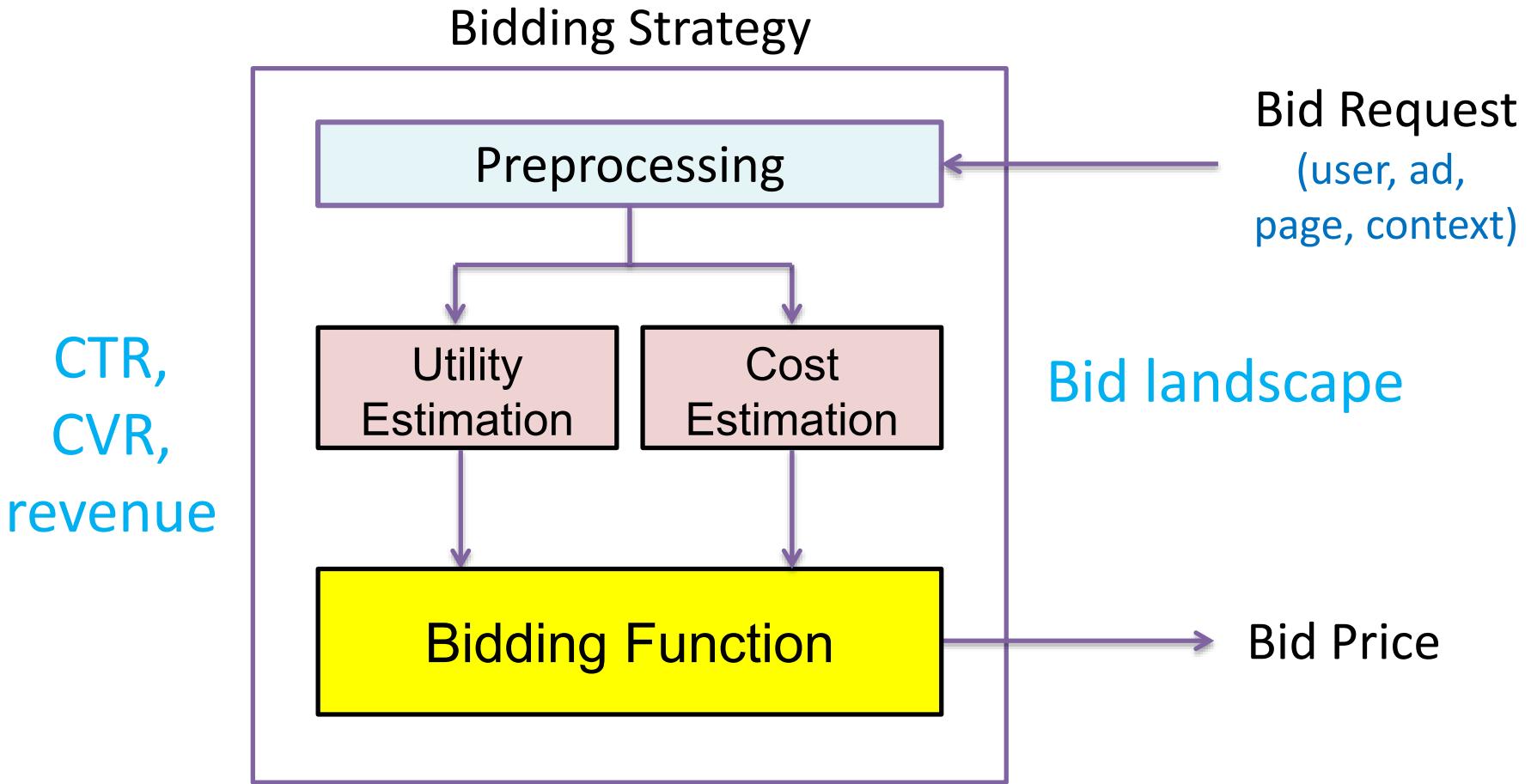
- Bid to optimise the KPI with budget constraint

$$\begin{array}{ll} \max_{\text{bidding strategy}} & \text{KPI} \\ \text{subject to} & \text{cost} \leq \text{budget} \end{array}$$

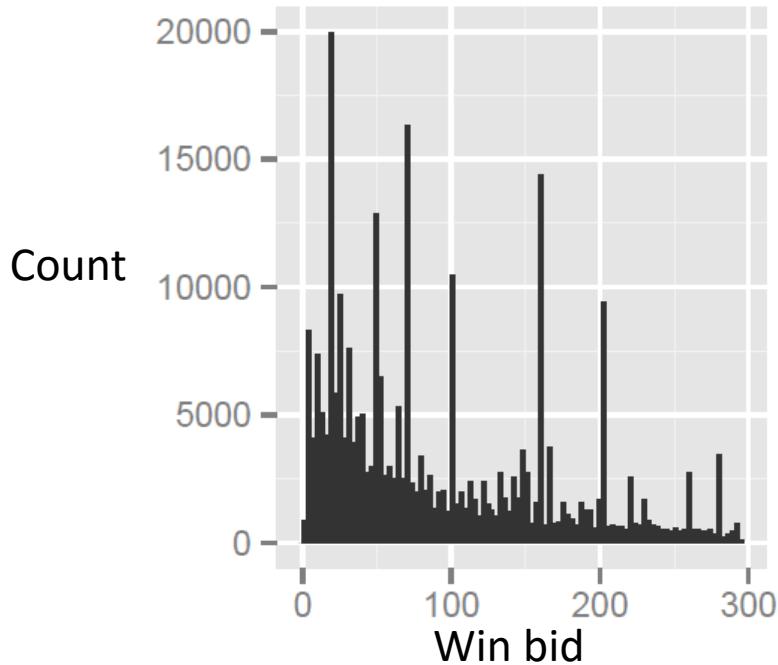
Bidding Strategy in Practice



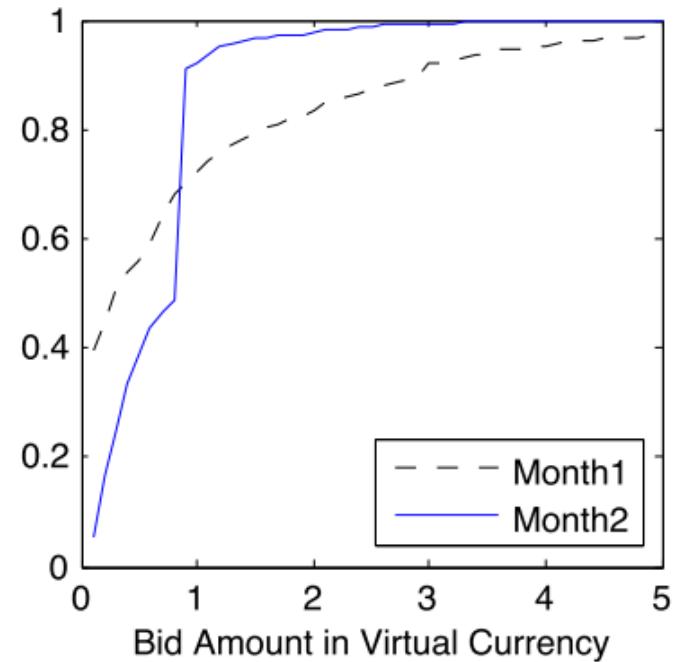
Bidding Strategy in Practice: A Quantitative Perspective



Bid Landscape Forecasting



Auction
Winning
Probability



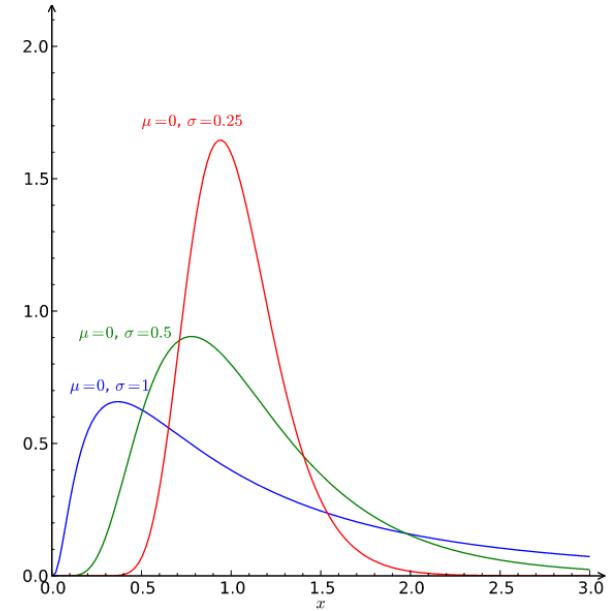
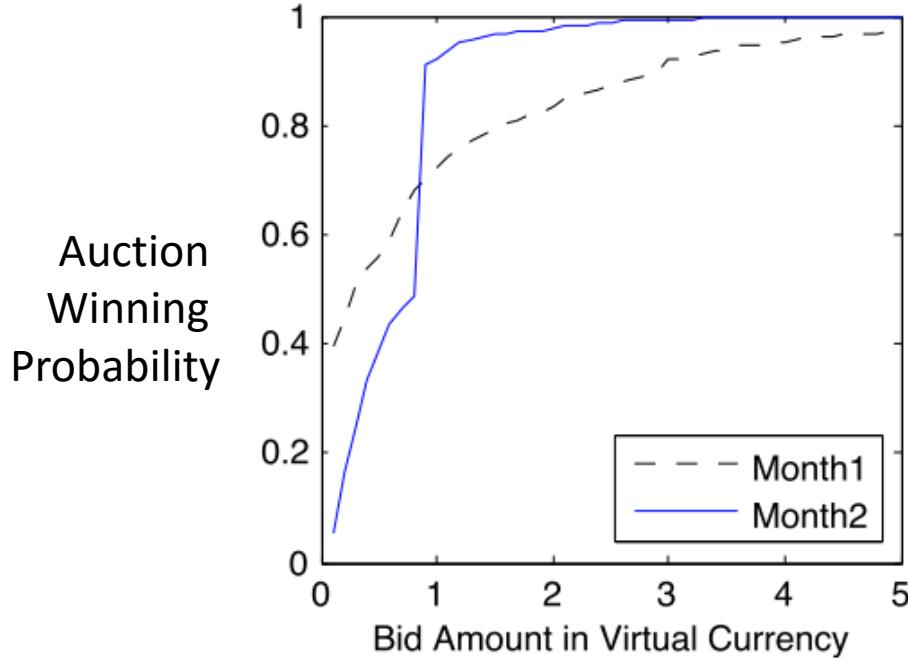
Win probability:

$$w(b) = \int_{z=0}^b p(z) dz$$

Expected cost:

$$c(b) = \frac{\int_{z=0}^b z p(z) dz}{\int_{z=0}^b p(z) dz}$$

Bid Landscape Forecasting



- Log-Normal Distribution

$$f_s(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x - \mu)^2}{2\sigma^2}}, x > 0$$

Bid Landscape Forecasting

- Price Prediction via Linear Regression

$$z = \boldsymbol{\beta}^T \mathbf{x} + \epsilon \quad \max_{\boldsymbol{\beta}} \sum_{i \in W} \log \phi\left(\frac{z_i - \boldsymbol{\beta}^T \mathbf{x}_i}{\sigma}\right)$$

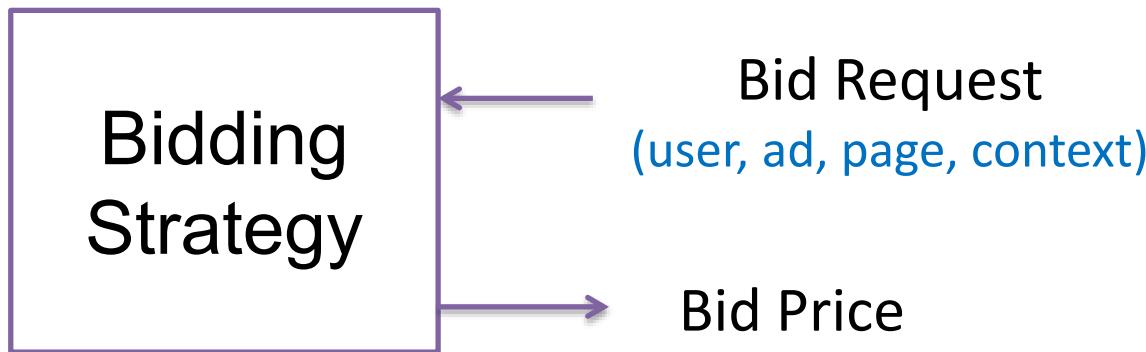
- Modelling censored data in lost bid requests

$$P(b_i < z_i) = \Phi\left(\frac{\boldsymbol{\beta}^T \mathbf{x}_i - b_i}{\sigma}\right)$$

$$\max_{\boldsymbol{\beta}} \sum_{i \in W} \log \phi\left(\frac{z_i - \boldsymbol{\beta}^T \mathbf{x}_i}{\sigma}\right) + \sum_{i \in L} \log \Phi\left(\frac{\boldsymbol{\beta}^T \mathbf{x}_i - b_i}{\sigma}\right)$$

Bidding Strategies

- How much to bid for each bid request?



- Bid to optimise the KPI with budget constraint

$$\begin{array}{ll} \max_{\text{bidding strategy}} & \text{KPI} \\ \text{subject to} & \text{cost} \leq \text{budget} \end{array}$$

Classic Second Price Auctions

- Single item, second price (i.e. pay market price)

Reward given a bid: $R(b) = \int_0^b (r - z)p(z)dz$

Optimal bid: $b^* = \max_b R(b)$

$$\frac{\partial R(b)}{\partial b} = (r - b)p(b)$$

$$\frac{\partial R(b)}{\partial b} = 0 \Rightarrow b^* = r \quad \text{Bid true value}$$

Truth-telling Bidding Strategies

- Truthful bidding in second-price auction
 - Bid the true value of the impression
 - Impression true value = $\begin{cases} \text{Value of click, if clicked} \\ 0, \text{ if not clicked} \end{cases}$
 - Averaged impression value = value of click * CTR
 - Truth-telling bidding:

$$\text{bid} = r_{\text{conv}} \times \text{CVR} \quad \text{or} \quad \text{bid} = r_{\text{click}} \times \text{CTR}$$

Truth-telling Bidding Strategies

$$\text{bid} = r_{\text{conv}} \times \text{CVR} \quad \text{or} \quad \text{bid} = r_{\text{click}} \times \text{CTR}$$

- Pros
 - Theoretic soundness
 - Easy implementation (very widely used)
- Cons
 - Not considering the constraints of
 - Campaign lifetime auction volume
 - Campaign budget
 - Case 1: \$1000 budget, 1 auction
 - Case 2: \$1 budget, 1000 auctions

Non-truthful Linear Bidding

- Non-truthful linear bidding

$$\text{bid} = \text{base_bid} \times \frac{\text{predicted_CTR}}{\text{base_CTR}}$$

- Tune `base_bid` parameter to maximise KPI
- Bid landscape, campaign volume and budget indirectly considered

$$\begin{array}{ll} \max_{\text{bidding strategy}} & \text{KPI} \\ \text{subject to} & \text{cost} \leq \text{budget} \end{array}$$

RTB Bidding Strategies

- Direct functional optimisation

$$b()_{\text{ORTB}} = \arg \max_{b()} N_T \int_{\theta} \theta w(b(\theta)) p_{\theta}(\theta) d\theta$$

winning function CTR

subject to $N_T \int_{\theta} b(\theta) w(b(\theta)) p_{\theta}(\theta) d\theta \leq B \leftarrow \text{budget}$

Est. volume bidding function cost upperbound

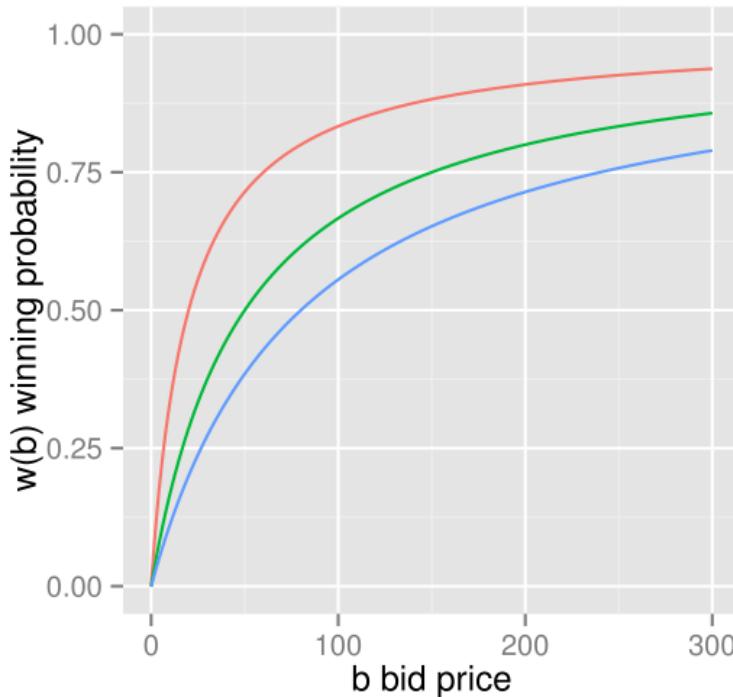
- Solution: Calculus of variations

$$\mathcal{L}(b(\theta), \lambda) = \int_{\theta} \theta w(b(\theta)) p_{\theta}(\theta) d\theta - \lambda \int_{\theta} b(\theta) w(b(\theta)) p_{\theta}(\theta) d\theta + \frac{\lambda B}{N_T}$$

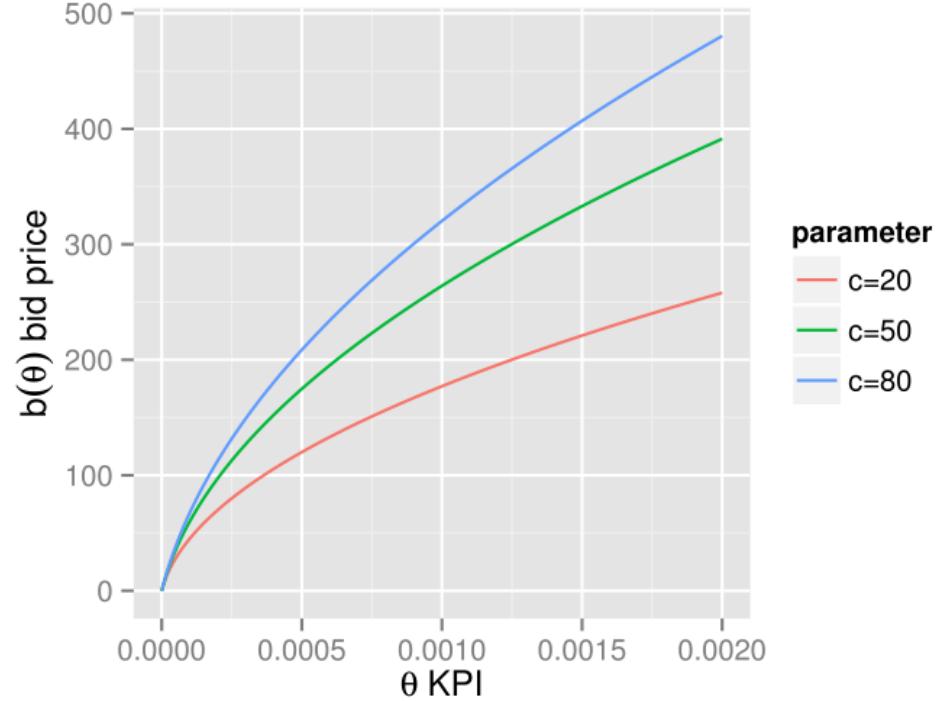
$\frac{\partial \mathcal{L}(b(\theta), \lambda)}{\partial b(\theta)} = 0 \quad \rightarrow \quad \boxed{\lambda w(b(\theta)) = [\theta - \lambda b(\theta)] \frac{\partial w(b(\theta))}{\partial b(\theta)}}$

[Zhang et al. Optimal real-time bidding for display advertising. KDD 14]

Optimal Bidding Strategy Solution



(a) Winning function 1.



(b) Bidding function 1.

$$w(b(\theta)) = \frac{b(\theta)}{c + b(\theta)} \rightarrow b_{\text{ORTB1}}(\theta) = \sqrt{\frac{c}{\lambda} \theta + c^2} - c$$

Bidding in Multi-Touch Attribution Mechanism

- Current bidding strategy
 - Driven by last-touch attribution $b(\text{CVR})$

$$\text{bid} = r_{\text{conv}} \times \text{CVR}$$

- A new bidding strategy
 - Driven by multi-touch attribution

$$\text{bid} = r_{\text{conv}} \times \text{CVR} \times P(\text{attribution} | \text{conversion})$$

$$\Delta P = P(y|S, a) - P(y|S)$$

$$\text{bid} = \Delta P \times \text{base_bid}$$

Table of contents

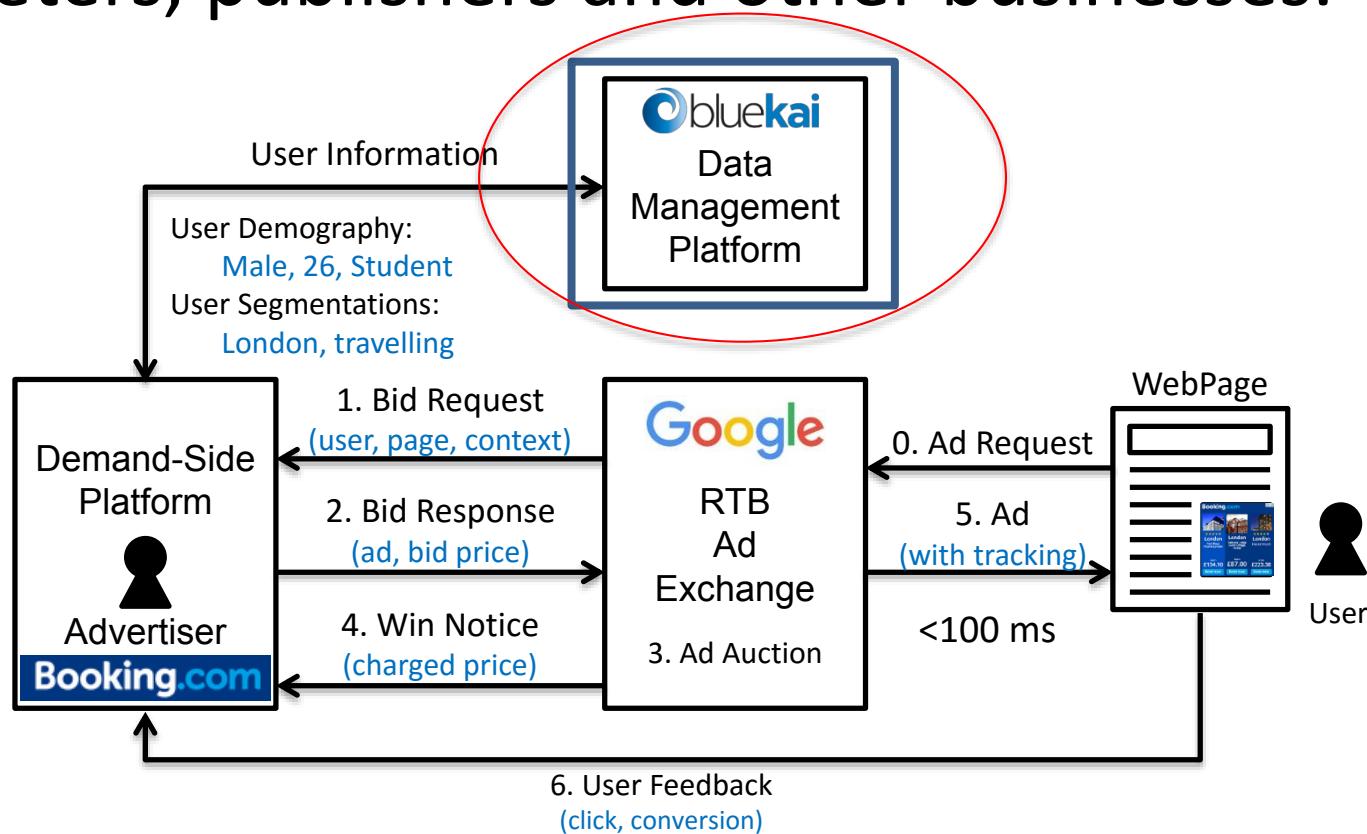
- RTB System
- Auction Mechanisms
- CTR Estimation
- Conversion Attribution
- Learning to Bid
- **Data Management Platform (DMP) techniques**
- Floor price optimisation
- Fighting against fraud

DMP Summary

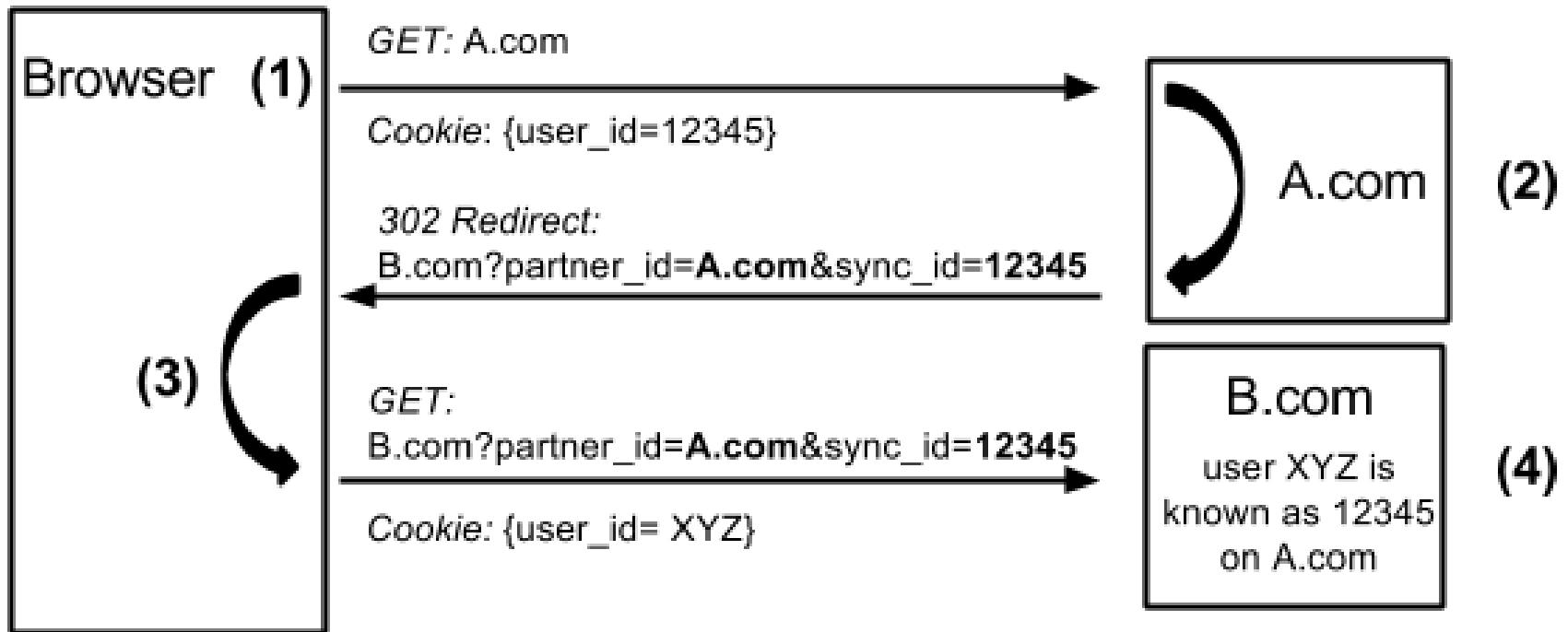
- What is data management platform
- Cook sync
- Browser fingerprinting
- CF and Lookalike model

What is DMP (Data Management Platform)

- A data warehouse that stores, merges, and sorts, and labels it out in a way that's useful for marketers, publishers and other businesses.



Cookie sync: merging audience data



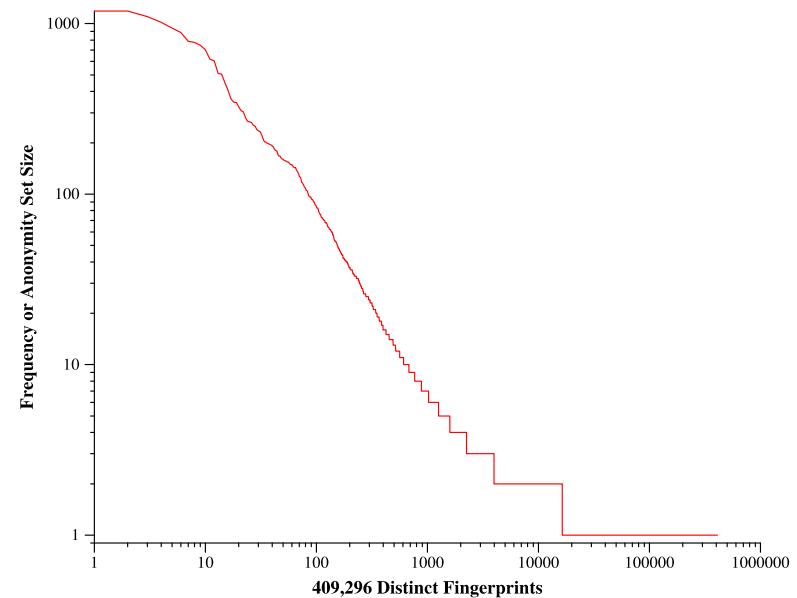
When a user visits a site (e.g. ABC.com) including A.com as a third-party tracker.

- (1) The browser makes a request to A.com, and included in this request is the tracking cookie set by A.com.
- (2) A.com retrieves its tracking ID from the cookie, and redirects the browser to B.com, encoding the tracking ID into the URL.
- (3) The browser then makes a request to B.com, which includes the full URL A.com redirected to as well as B.com's tracking cookie.
- (4) B.com can then link its ID for the user to A.com's ID for the user2

Browser fingerprinting

- A device fingerprint or browser fingerprint is information collected about the remote computing device for the purpose of identifying the user
- Fingerprints can be used to fully or partially identify individual users or devices even when cookies are turned off.

94.2% of browsers with Flash or Java were unique in a study

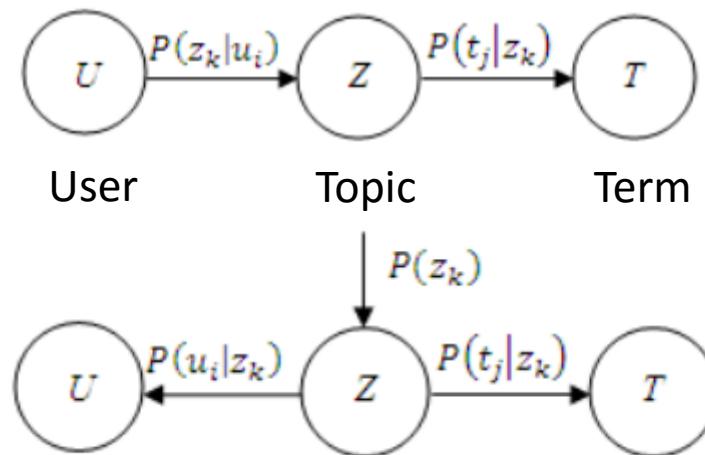


Eckersley, Peter. "How unique is your web browser?." Privacy Enhancing Technologies. Springer Berlin Heidelberg, 2010.

Acar, Gunes, et al. "The web never forgets: Persistent tracking mechanisms in the wild." Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2014.

User segmentation and Behavioural Targeting

- Behavioural targeting helps online advertising
- From user – documents to user – topics
 - Latent Semantic Analysis / Latent Dirichlet Allocation

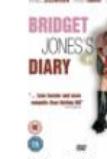


J Yan, et al., How much can behavioral targeting help online advertising? WWW 2009

X Wu, et al., Probabilistic latent semantic user segmentation for behavioral targeted advertising, Intelligence for Advertising 2009

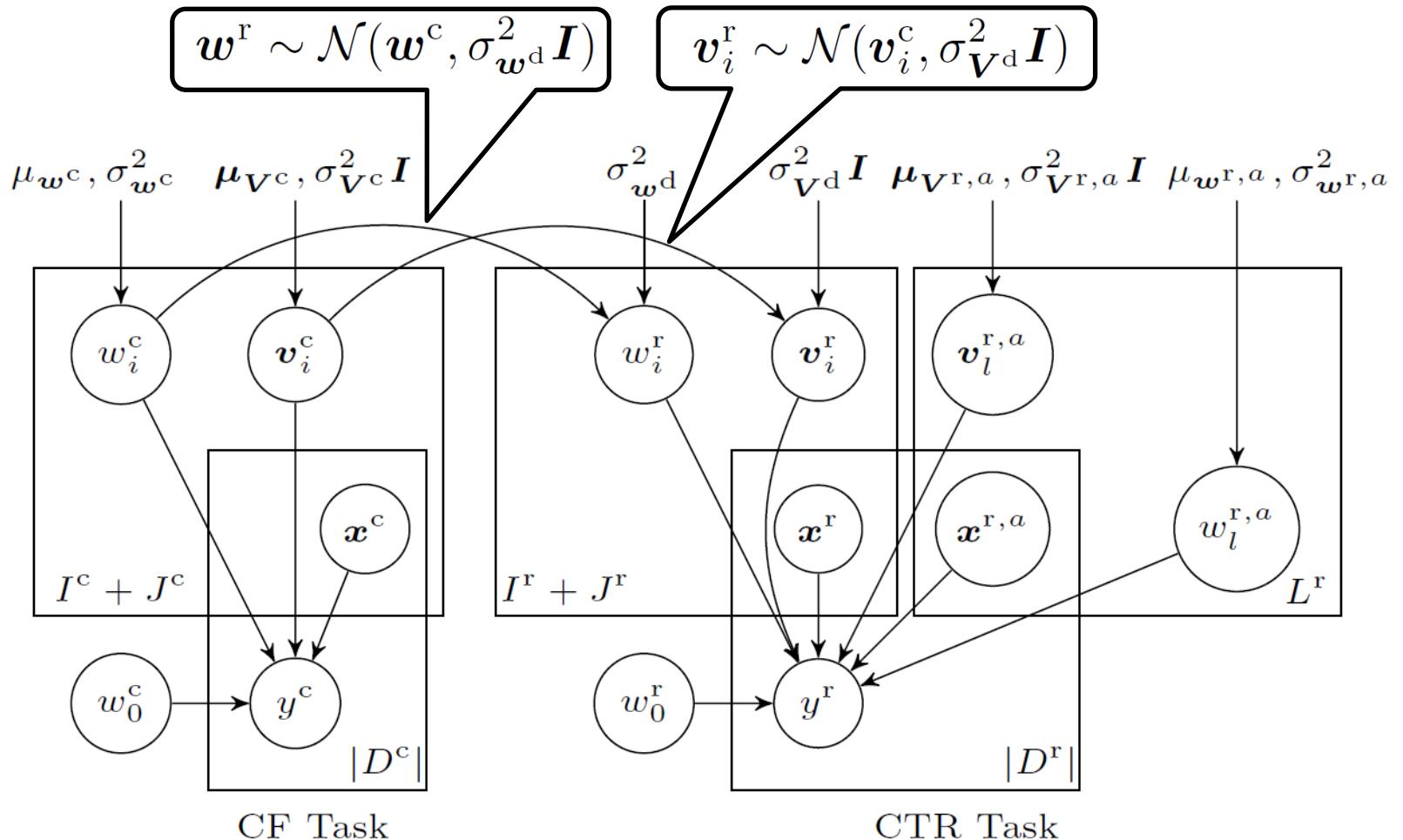
Lookalike modelling

- Lookalike modeling: finding new people who behave like current customers (converted)

	 Die Hard	 Mission: Impossible	 GoldenEye	 Casino Royale	 Titanic	 Notting Hill	 Bridget Jones's Diary	 Love Actually
► Boris	★★★★★	★★★★★	★★★★★			★★★★★		?
► Dave		★★★★★	★★★★★	★★★★★			★★★★★	
► Will		★★★★★			★★★★★	★★★★★	★★★★★	★★★★★
► George	★★★★★	★★★★★	★★★★★	★★★★★				★★★★★

Transferred lookalike

Using web browsing data, which is largely available, to infer the ad clicks

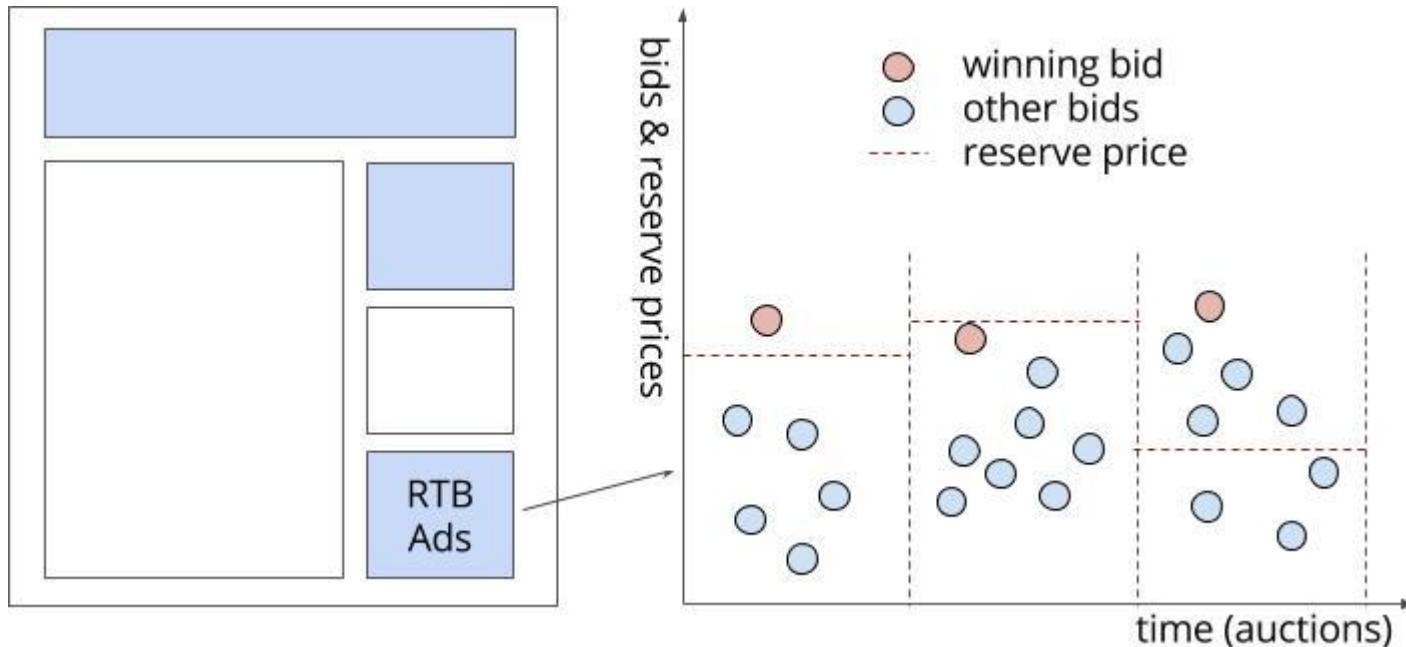


Zhang, Weinan, Lingxi Chen, and Jun Wang. "Implicit Look-alike Modelling in Display Ads: Transfer Collaborative Filtering to CTR Estimation." ECIR (2016). In Wednesday Information Filtering Track

Table of contents

- RTB System
- Auction Mechanisms
- CTR Estimation
- Conversion Attribution
- Learning to Bid
- Data Management Platform (DMP) techniques
- **Floor price optimisation**
- Fighting against fraud

Reserve price optimisation



The task:

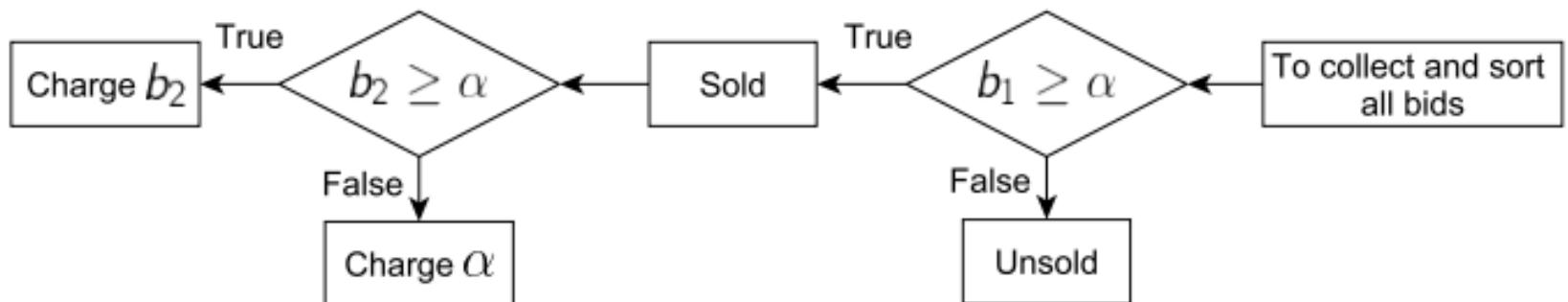
- To find the optimal reserve prices

The challenge:

- Practical constraints v.s common assumptions (bids' distribution, bidding private values, etc.)

Why

- Suppose it is second price auction
 - Normal case: $b_2 \geq \alpha$
 - Preferable case: $b_1 \geq \alpha > b_2$ (it increases the revenue)
 - Undesirable case: $\alpha > b_1$ (but there is risk)



An example

- Suppose: two bidders, private values drawn from Uniform[0, 1]
- Without a reserve price (or $a = 0$), the payoff r is:

$$r = E[\min(b_1, b_2)] = 0.33$$

- With $a = 0.2$:

$$r = E[\min(b_1, b_2) | b_1 > 0.2, b_2 > 0.2] + 0.32 \times 0.2 = 0.36$$

- With $a = 0.5$:

$$r = E[\min(b_1, b_2) | b_1 > 0.5, b_2 > 0.5] + 0.5 \times 0.5 = 0.42$$

- With $a = 0.6$:

$$r = \underline{E[\min(b_1, b_2) | b_1 > 0.6, b_2 > 0.6]} + \underline{(0.6 \times 0.4) \times 2 \times 0.6} = 0.405$$



Paying the second highest price Paying the reserve price

The optimal auction theory

- In the second price auctions, advertisers bid their private values $[b_1, \dots, b_K]$
$$F(\mathbf{b}) = F_1(b_1) \times \cdots \times F_K(b_K)$$
- Private values -> Bids' distributions
 - Uniform
 - Log-normal
- The publisher also has a private value V_p
- The optimal reserve price is given by: $\alpha - \frac{1 - F(\mathbf{b})}{F'(\mathbf{b})} - V_p = 0$

Results from a field experiment

- On Yahoo! Sponsored search
- Using the Optimal Auction Theory

Table 7: Restricted sample (optimal reserve price < 20¢)

Variable	Value	t-statistic	p-value
Number of keywords (T – treatment group)	222,249		
Number of keywords (C – control group)	11,615		
(Mean change in depth in T)–(mean change in depth in C)	-0.8612	-60.29	< 0.0001
(Mean change in revenue in T)–(mean change in revenue in C)	-11.88%	-2.45	0.0144
Estimated impact of reserve prices on revenues	-9.19%	-11.1	< 0.0001

Mixed results

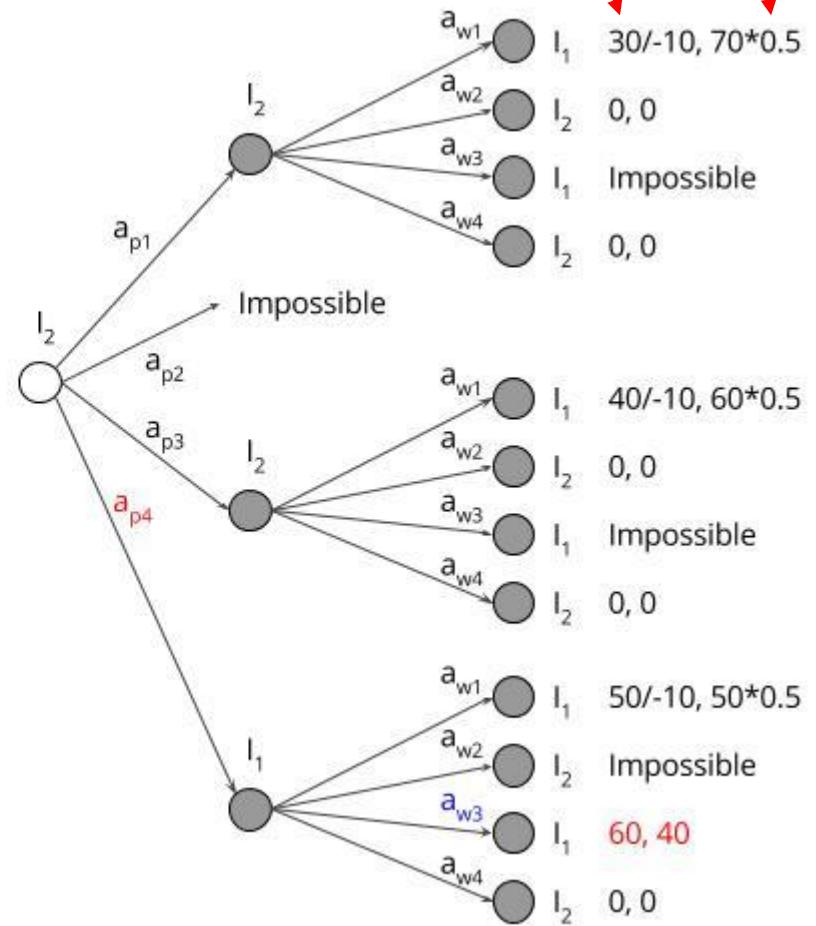
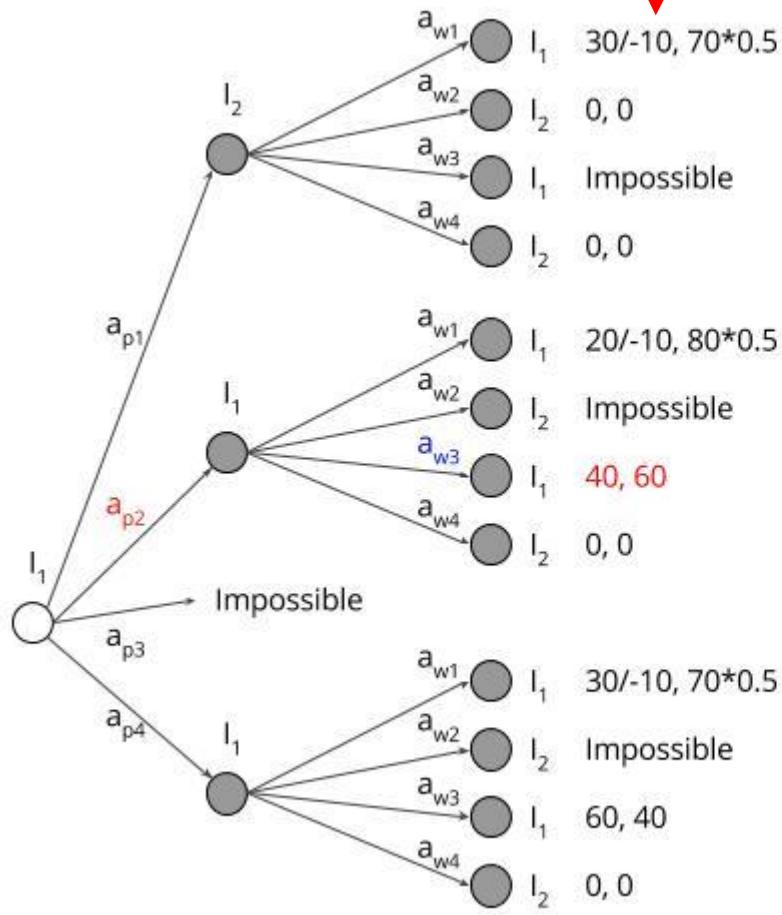
Table 8: Restricted sample (optimal reserve price \geq 20¢)

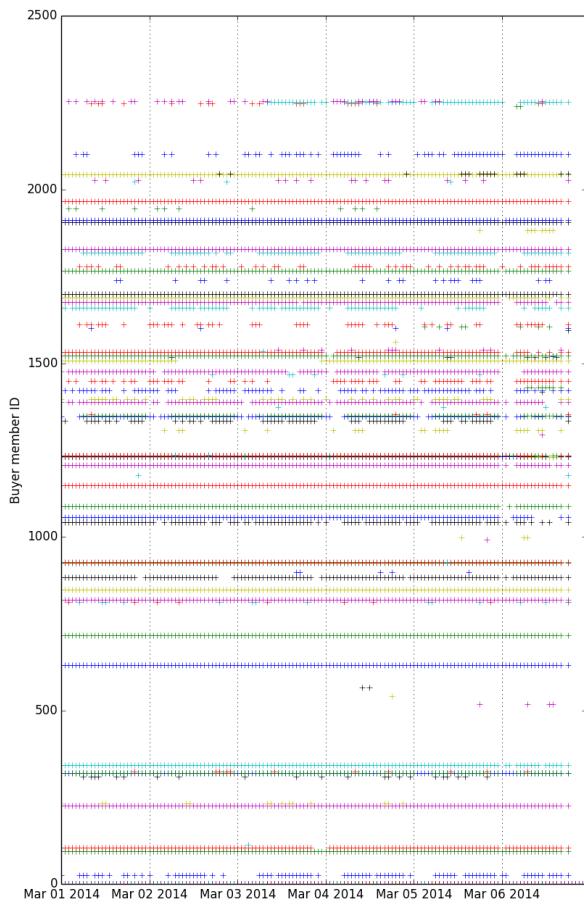
Variable	Value	t-statistic	p-value
Number of keywords (T – treatment group)	216,383		
Number of keywords (C – control group)	11,401		
(Mean change in depth in T)–(mean change in depth in C)	-0.9664	-55.09	< 0.0001
(Mean change in revenue in T)–(mean change in revenue in C)	14.59%	1.79	0.0736
Estimated impact of reserve prices on revenues	3.80%	5.41	< 0.0001

1) Expected payoff of advertiser, publisher

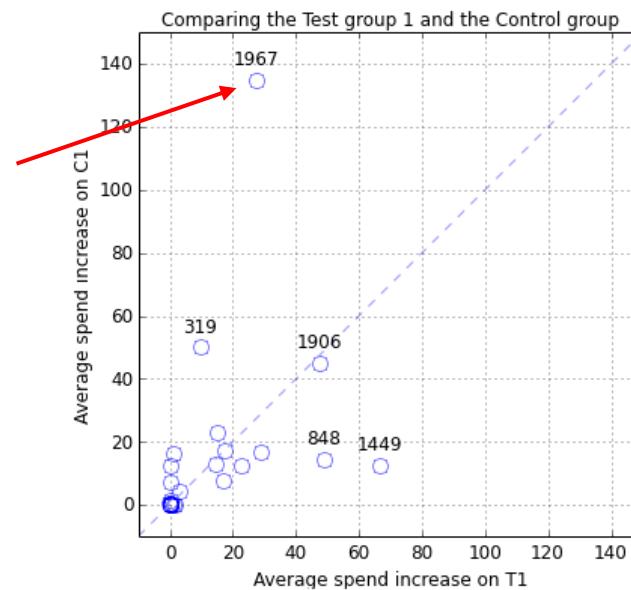
2) Payoff for the advertiser could be negative if one has been bidding the max price
 $(a_{w1}$: to increase b_1 so that $b_1 \geq \alpha$)

3) One won't do that, so discounted publisher's payoff

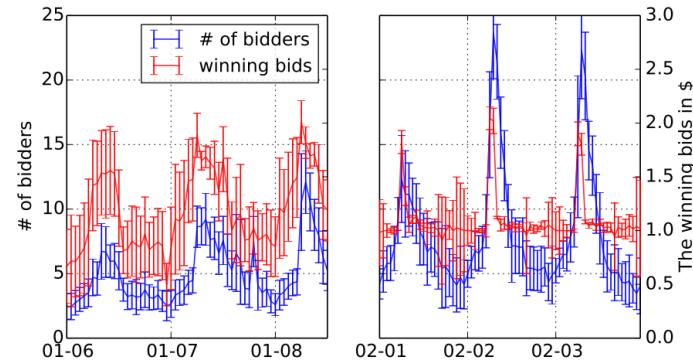




An outlier
(Triggered by some random action)



The unchanged budget allocation



The unchanged bidding pattern

Table of contents

- RTB System
- Auction Mechanisms
- CTR Estimation
- Conversion Attribution
- Learning to Bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud

Fighting publisher fraud

- Non intentional traffic (NIT) / Non human traffic
 - Web scrapers / crawlers
 - Hacking tools
 - Botnet
 - Much of the spurious traffic is created by human but without users' knowledge

A Serious Problem

We realized this by testing out a buying platform in Atlas last year. During that test, we plugged into a number of the usual exchanges and bought across several formats. There were two major takeaways:

1. We were able to deliver ads to real people with unprecedented accuracy, but came up against many bad ads and fraud (like bots). While we were fortunately able to root out the bad actors and only buy quality ads, we were amazed by the volume of valueless inventory.
2. Only two ad formats delivered significant value: native & video.

Based on those findings, we began to dig into the ads that came through LiveRail. And when we saw the same thing, we immediately shut off the low quality ads. In fact, we removed over 75% of the volume coming from our exchange by turning off publishers circulating bad inventory into LiveRail. We knew that in good conscience, we couldn't sell what Atlas and our people-based measurement told us was valueless. Unfortunately, those ads were almost certainly dumped into another low-quality exchange where all of them were most likely purchased.

Dave Jakubowski, Head of Ad Tech, Facebook, March 2016

The Old Fashion Way

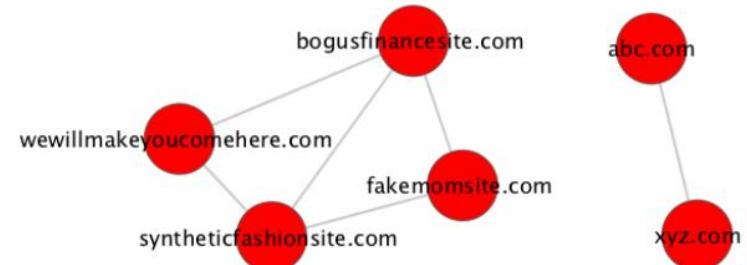
- Put the police on the street
 - Manually eyeball the webpage
 - Verify the address on the Google map
- Follow how the money flows
- This approach just can't scale and is not sustainable

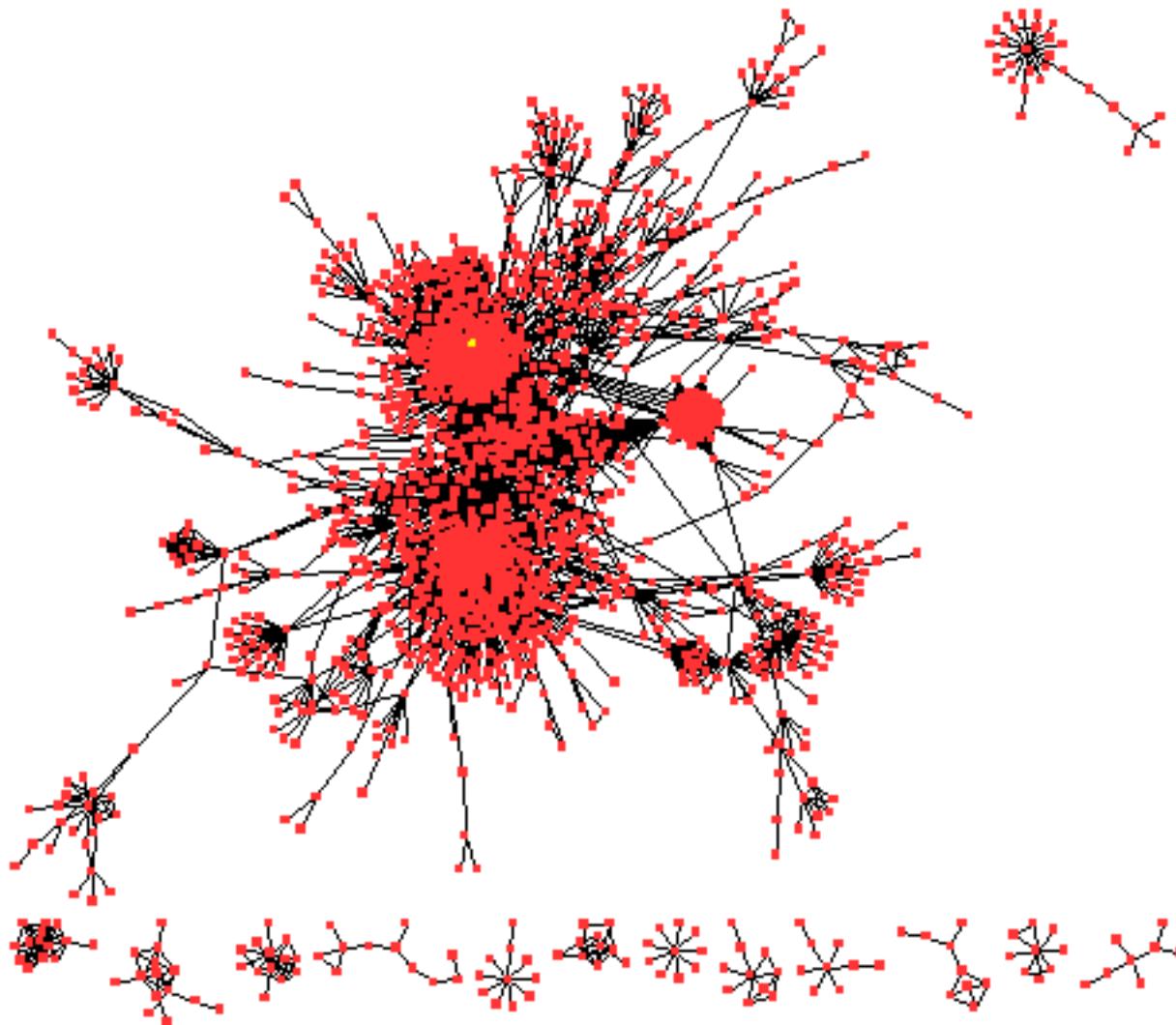
Possible Solutions

- Rules
- Anomaly detection
- Classification algorithm
 - Tricky to obtain negative samples
- Clustering algorithm
 - Bots could display dramatically different behavior
- Content Analysis
 - Fraudulent websites often scrape content from each other or legit websites

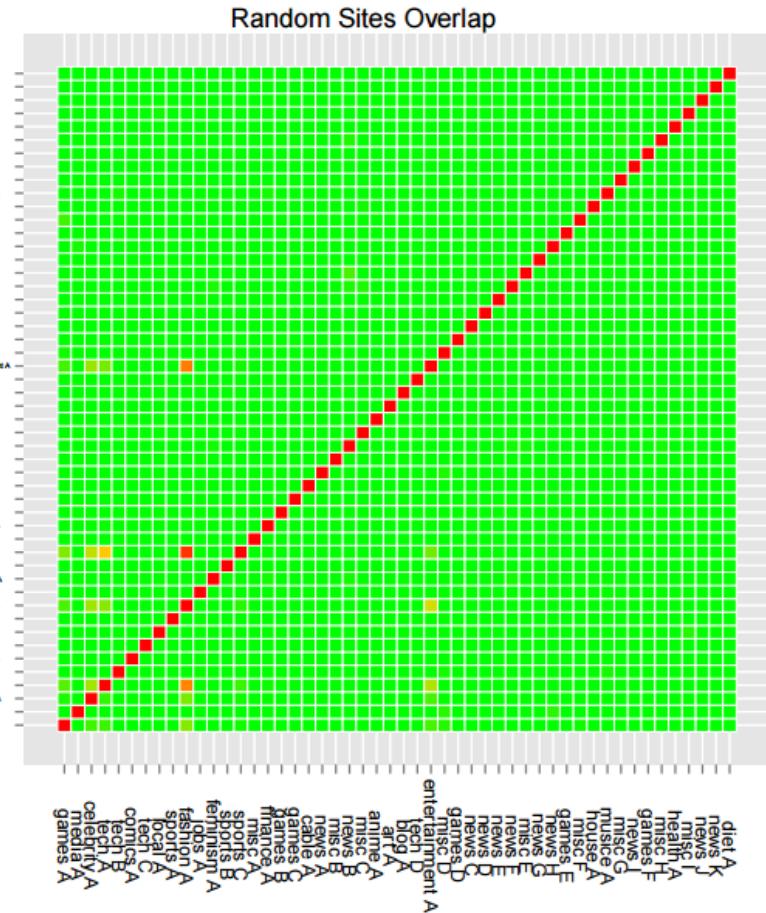
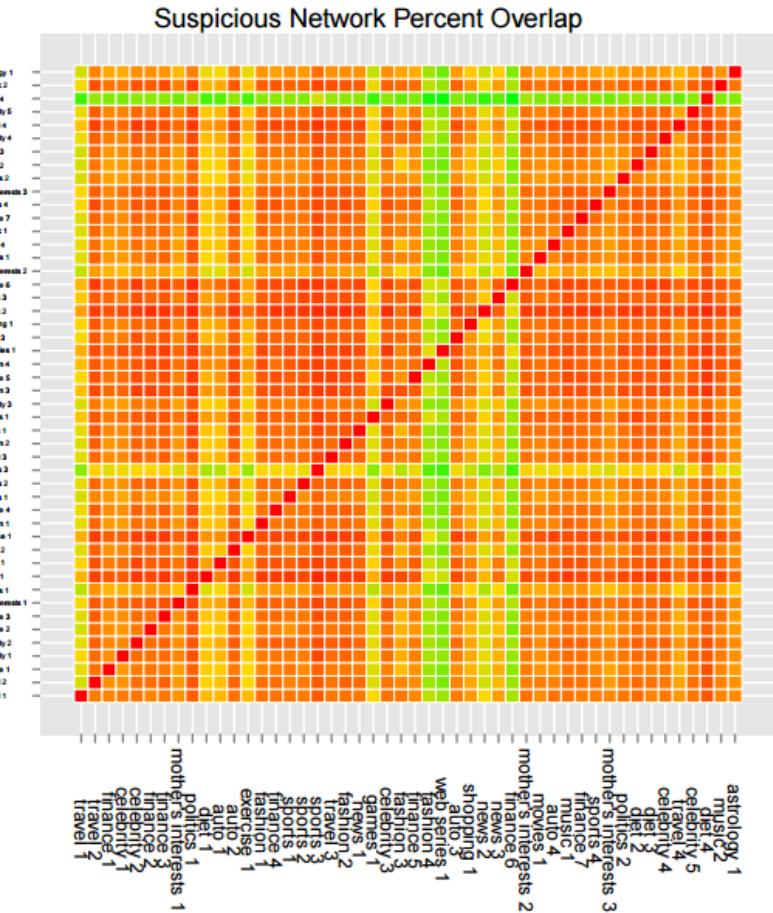
Co-Visitation Networks

- Key observation:
 - Even the major sites only share at most 20% cookieID within a few hours, let alone those long tail sites.
- Define a graph:
 - Node: site
 - Weighted edge: user overlap ratio of two sites
- Cluster this weighted undirected graph
- Fraud: big cluster with long tail sites





December 2011 Co-visitation Network where an edge indicates at least 50% overlap between the browsers of both websites



Real-Time Bidding based Display Advertising: Mechanisms and Algorithms

Thank You

- RTB system
- Auction mechanisms
- User response estimation
- Conversion attribution
- Learning to bid
- Data Management Platform (DMP) techniques
- Floor price optimisation
- Fighting against fraud