CIKM16 Tutorial

Learning, Prediction and Optimisation in RTB Display Advertising

Weinan Zhang, Shanghai Jiao Tong University
Jian Xu, TouchPal Inc.

http://www.optimalrtb.com/cikm16/

October 24, 2016, Indianapolis, United States

Speakers

- Weinan Zhang
  - Assistant Professor at Shanghai Jiao Tong University
  - Ph.D. from University College London 2016
  - Machine learning, data mining in computational advertising and recommender systems

- Jian Xu
  - Principal Data Scientist at TouchPal, Mountain View
  - Previous Senior Data Scientist and Senior Research Engineer at Yahoo! US
  - Data mining, machine learning, and computational advertising

Tutorial Materials

- Web site:

  http://www.optimalrtb.com/cikm16

- Supporting documents:
  - RTB monograph

    https://arxiv.org/abs/1610.03013
  - RTB paper list:

    https://github.com/wnzhang/rtb-papers

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Learning to bid
- Conversion attribution
- Pacing control
- Targeting and audience expansion
- Reserve price optimization

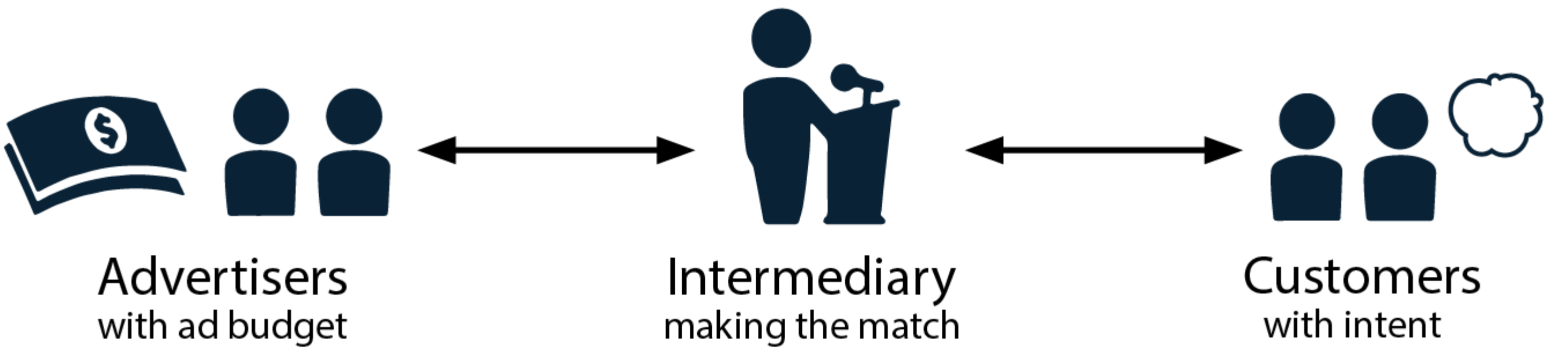
Weinan Zhang
90 min

30 min break

Jian Xu
90 min

Table of contents

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."

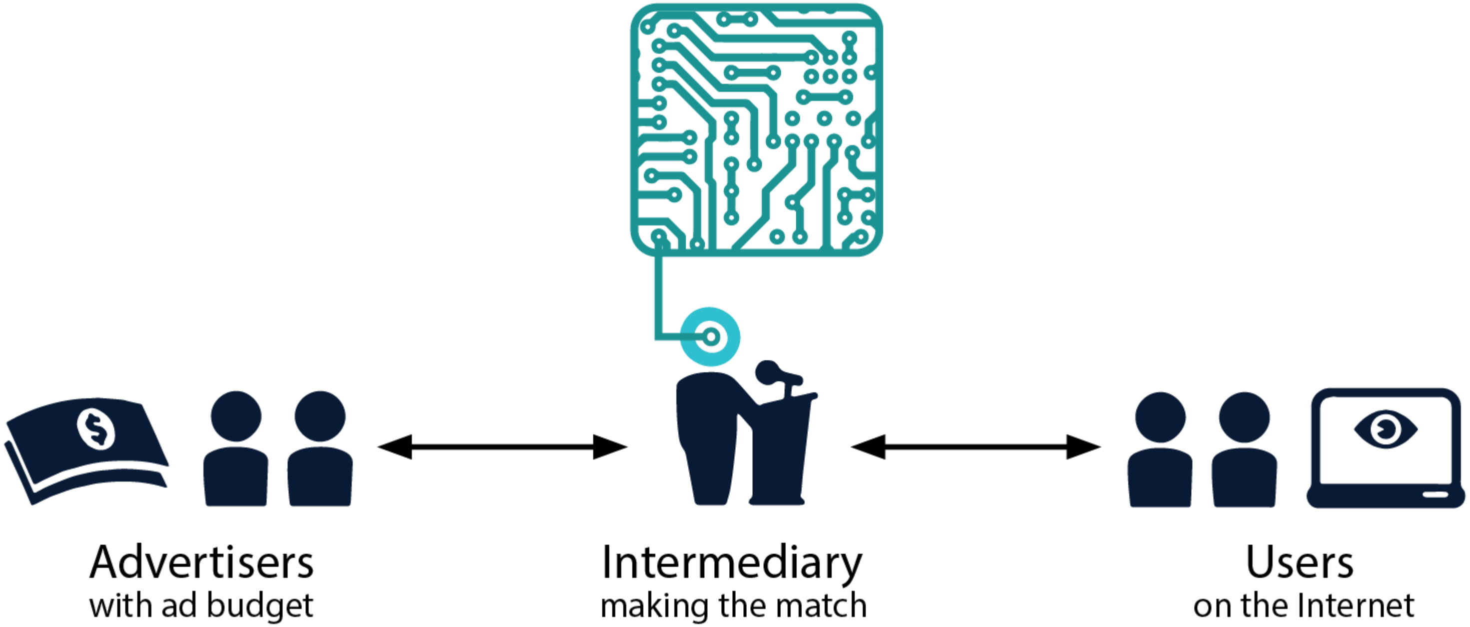
- *John Wanamaker*

(1838-1922)

Father of modern advertising and a pioneer in marketing

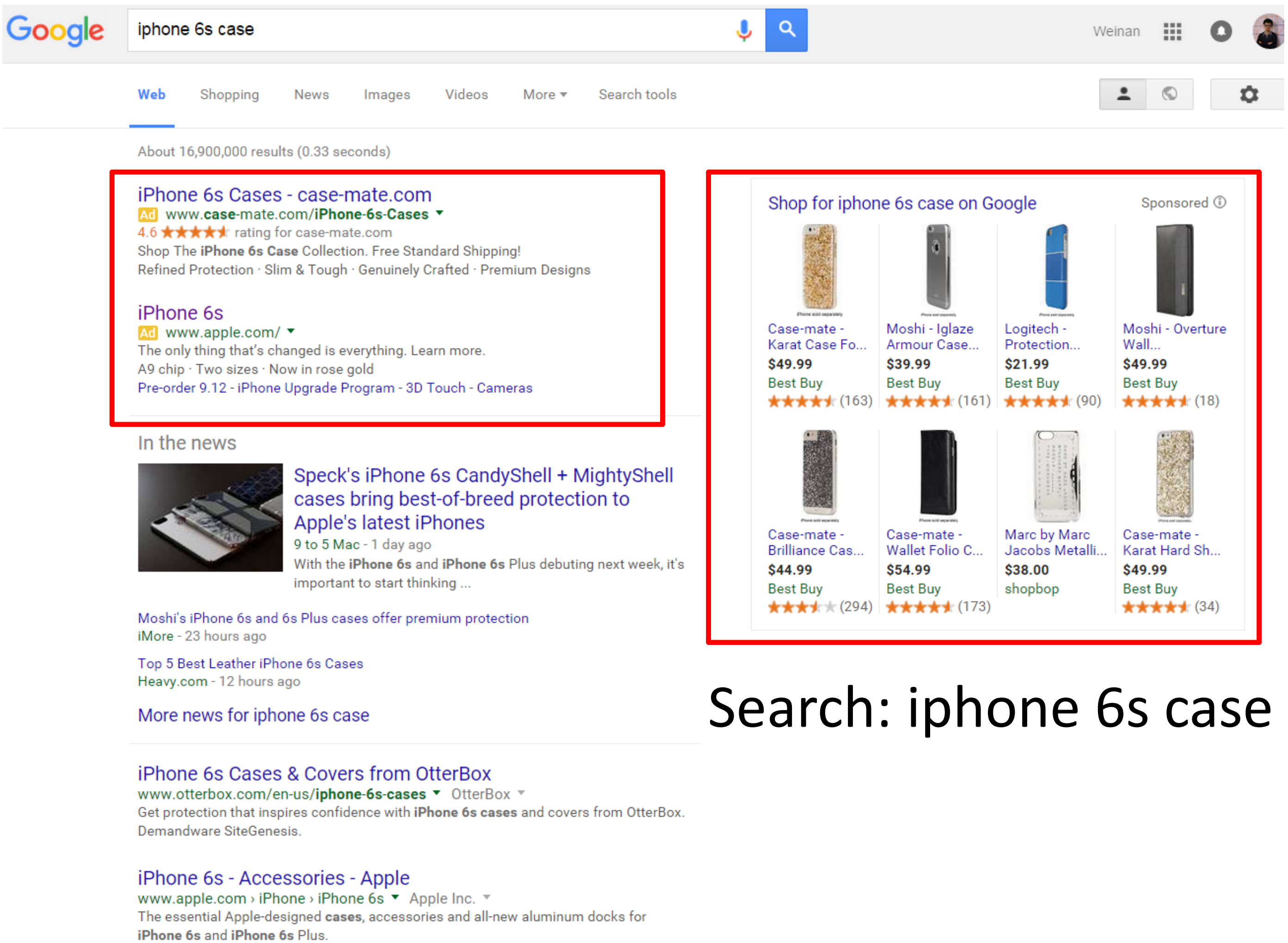
Wasteful Traditional Advertising

Computational Advertising



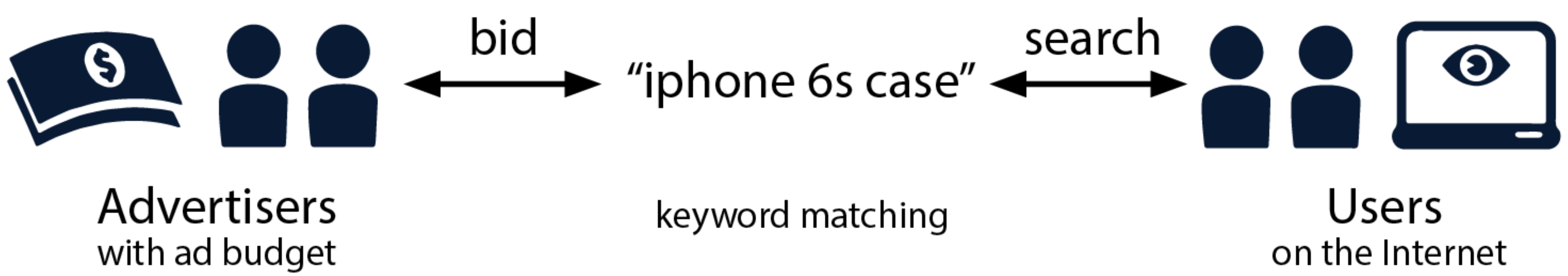
- Design **algorithms** to make the best match between the advertisers and Internet users with economic constraints

Sponsored Search



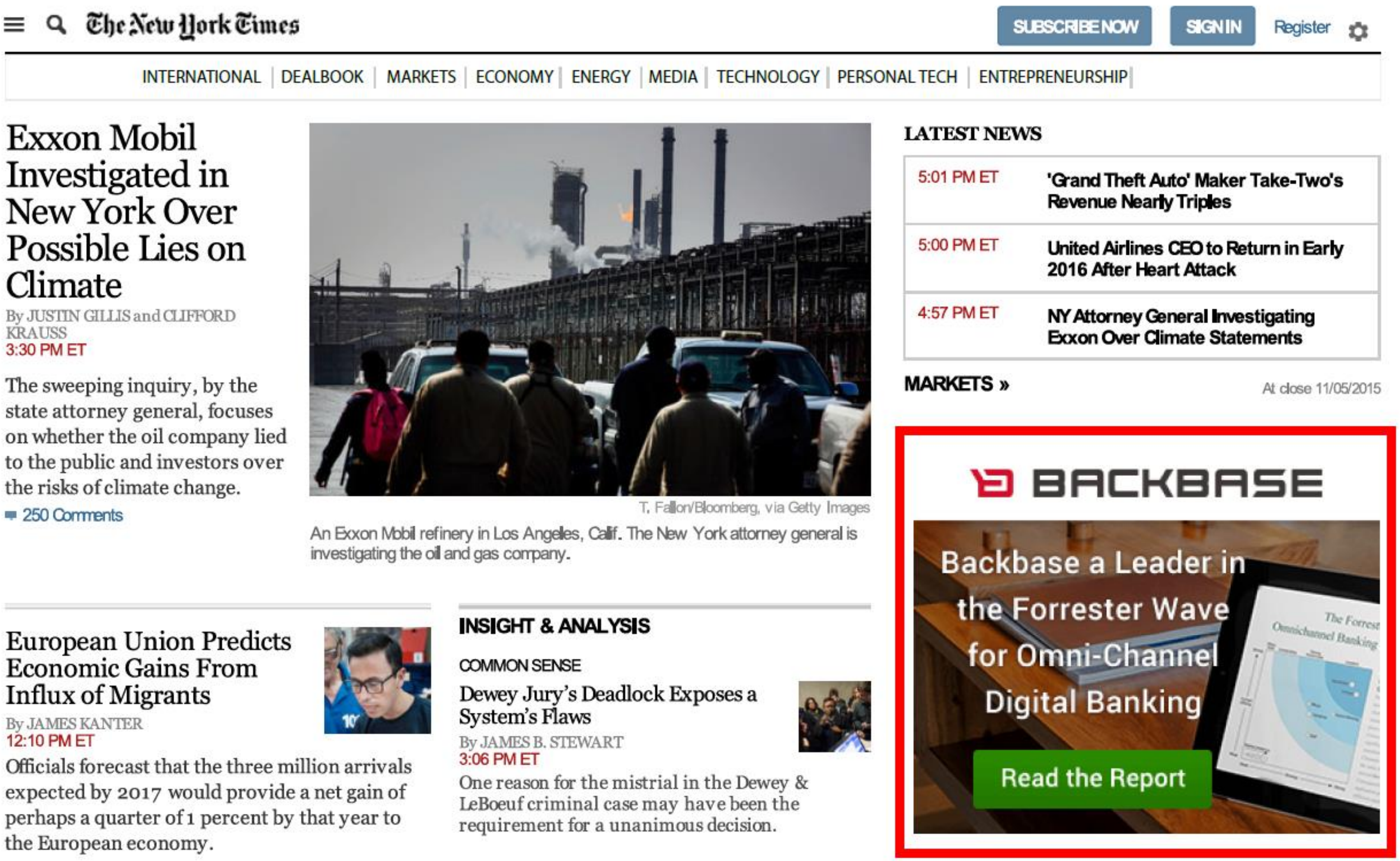
Search: iphone 6s case

Sponsored Search



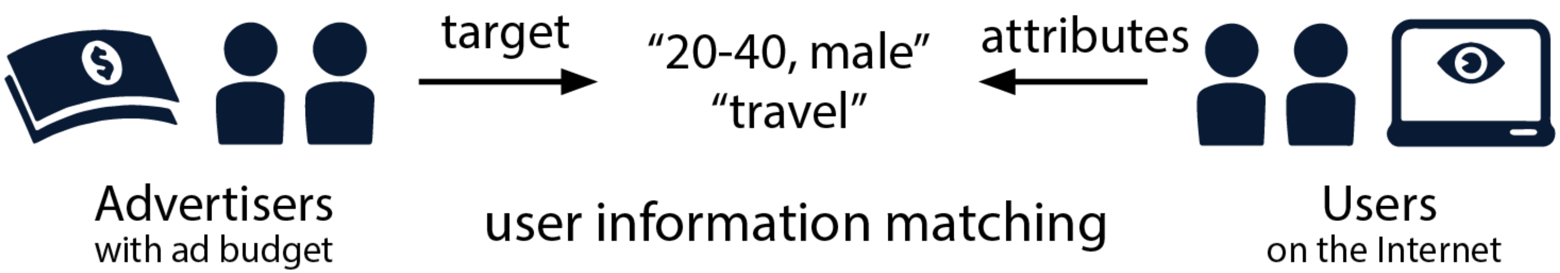
- Advertiser sets a bid price for the keyword
- User searches the keyword
- Search engine hosts the auction to ranking the ads

Display Advertising



http://www.nytimes.com/

Display Advertising



- Advertiser targets a segment of users
- Intermediary matches users and ads by user information

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 billon impressions
Finance	New York Stock Exchange	12 billion shares
	Shanghai Stock Exchange	14 billion shares

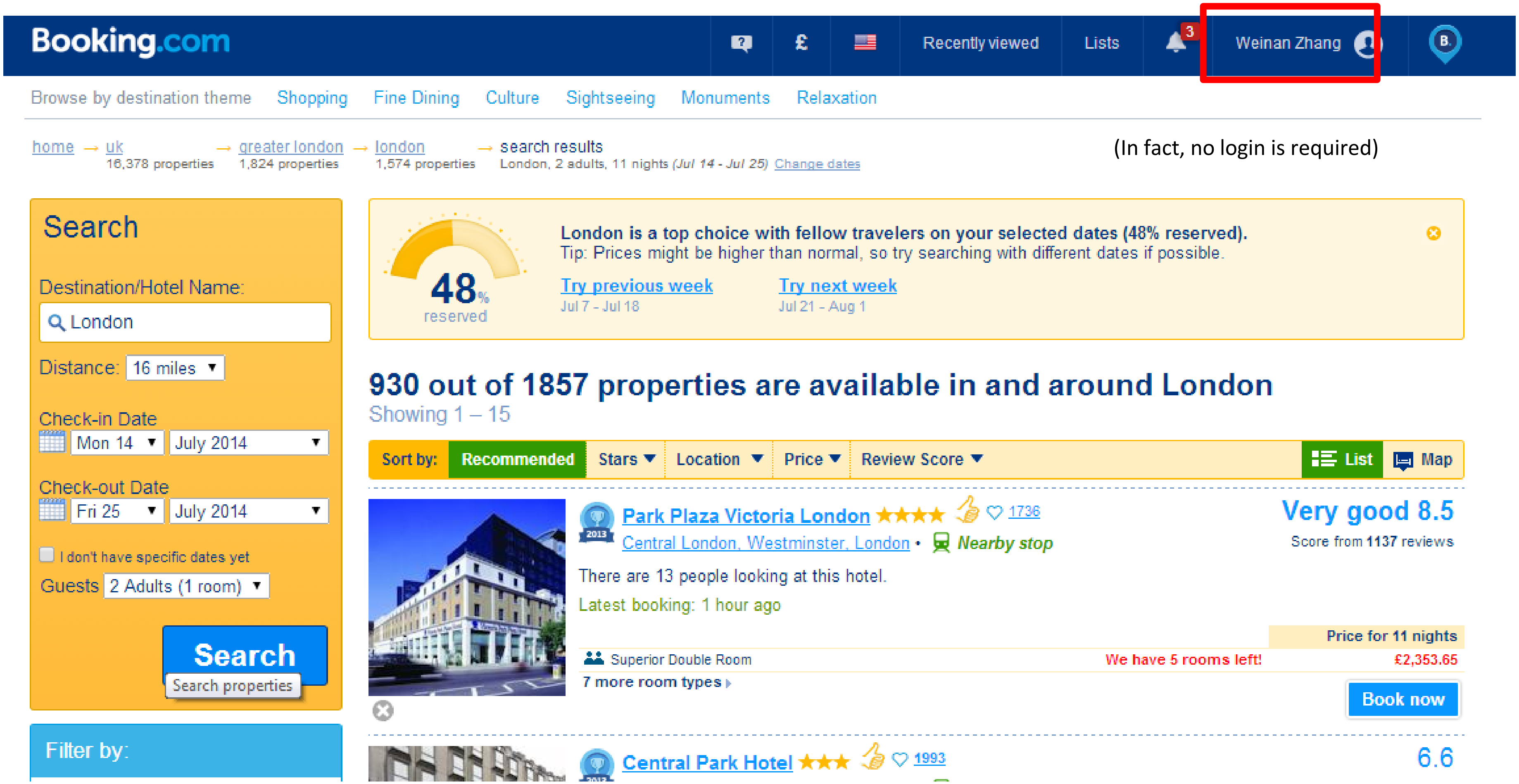
	Query per second
Turn DSP	1.6 million
Google	40,000 search

[Shen, Jianqiang, et al. "From 0.5 Million to 2.5 Million: Efficiently Scaling up Real-Time Bidding." Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015.]

Suppose a student regularly reads articles on emarketer.com

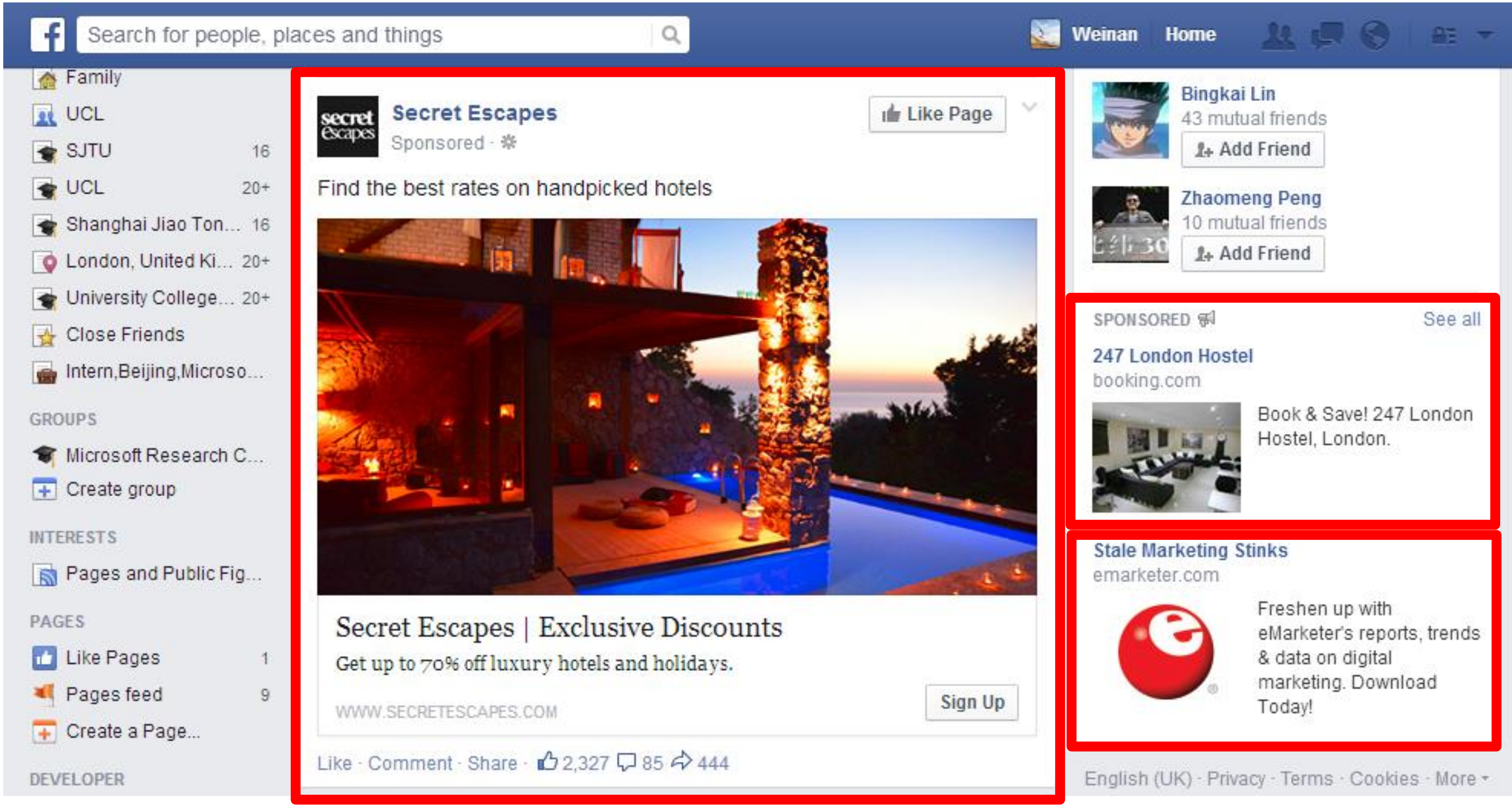
Content-related ads

He recently checked the London hotels

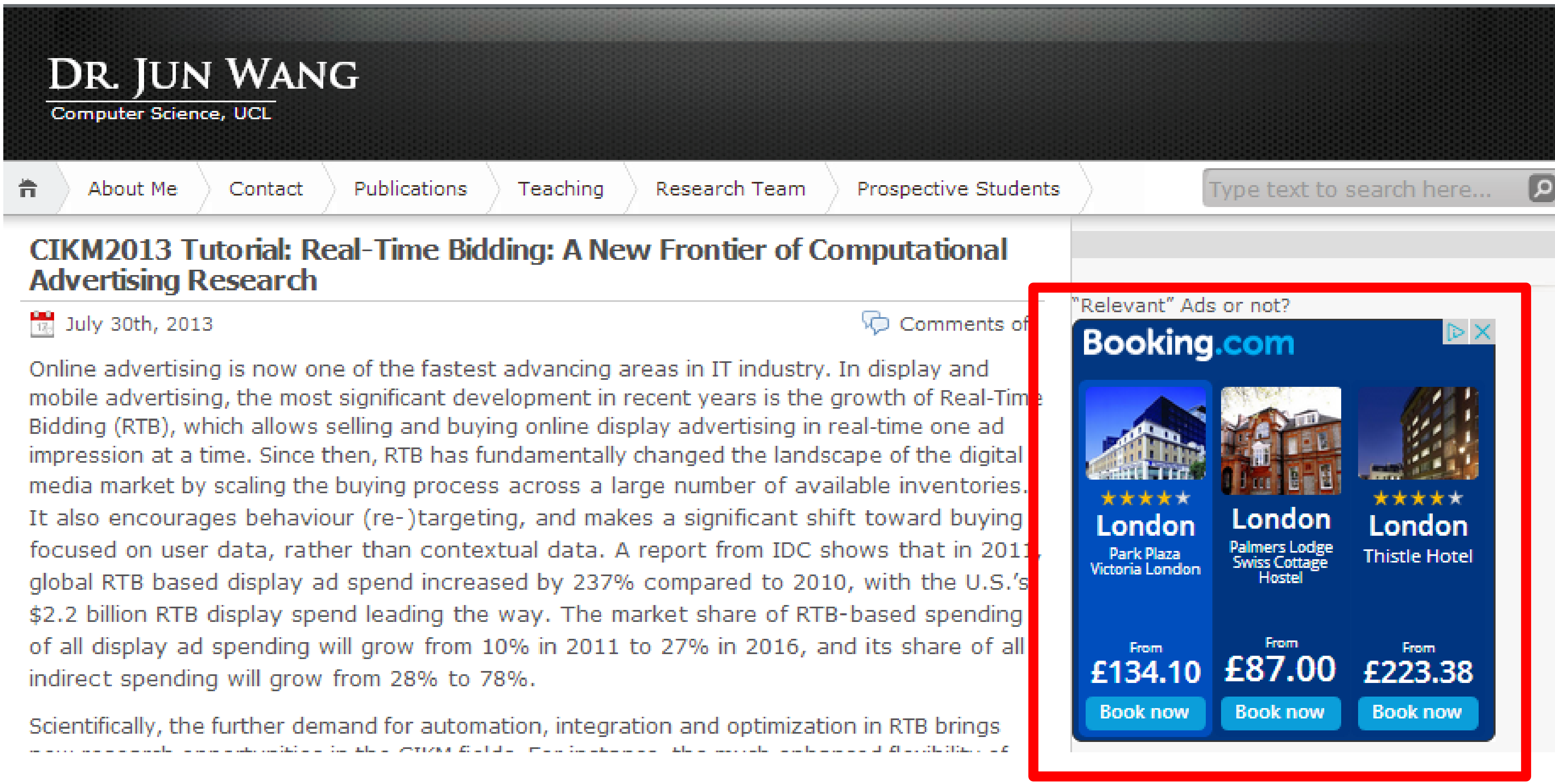


(In fact, no login is required)

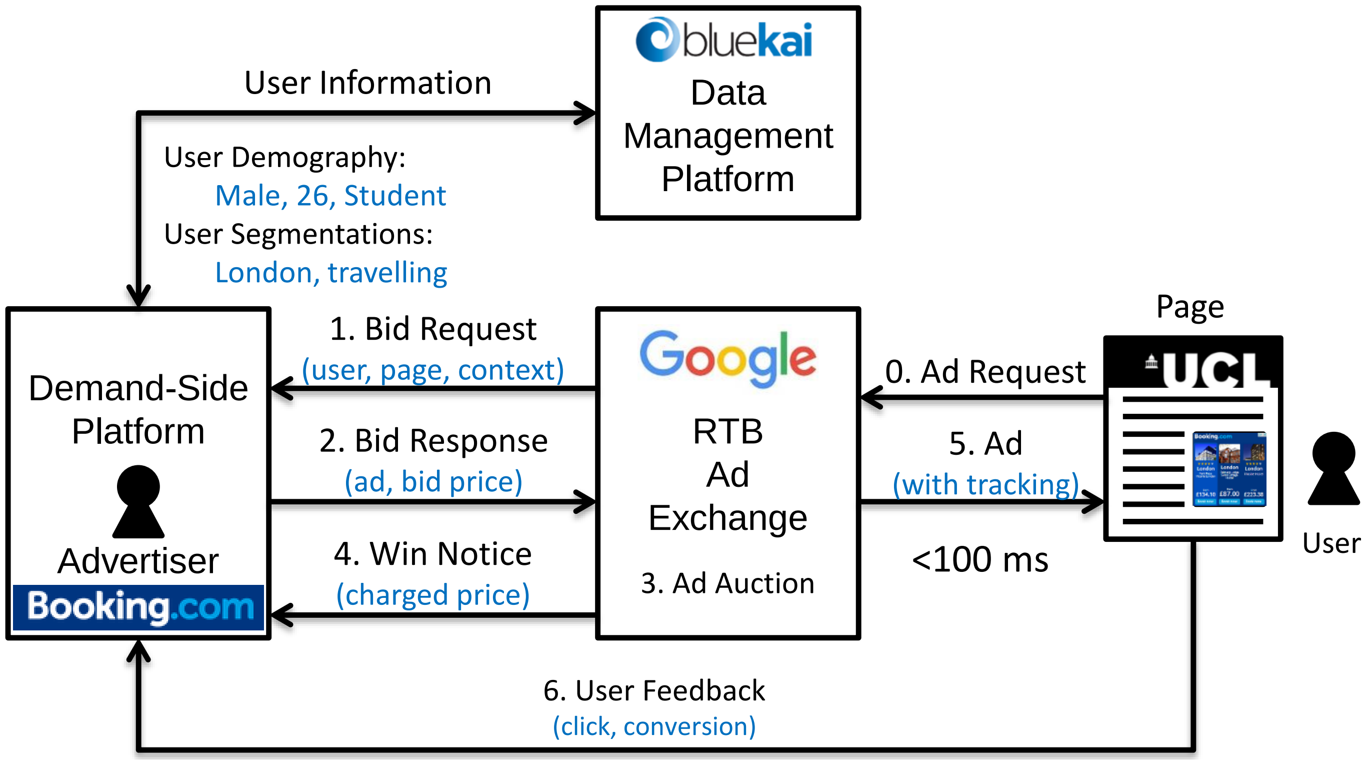
Relevant ads on facebook.com



Even on supervisor's homepage!
(User targeting dominates the context)



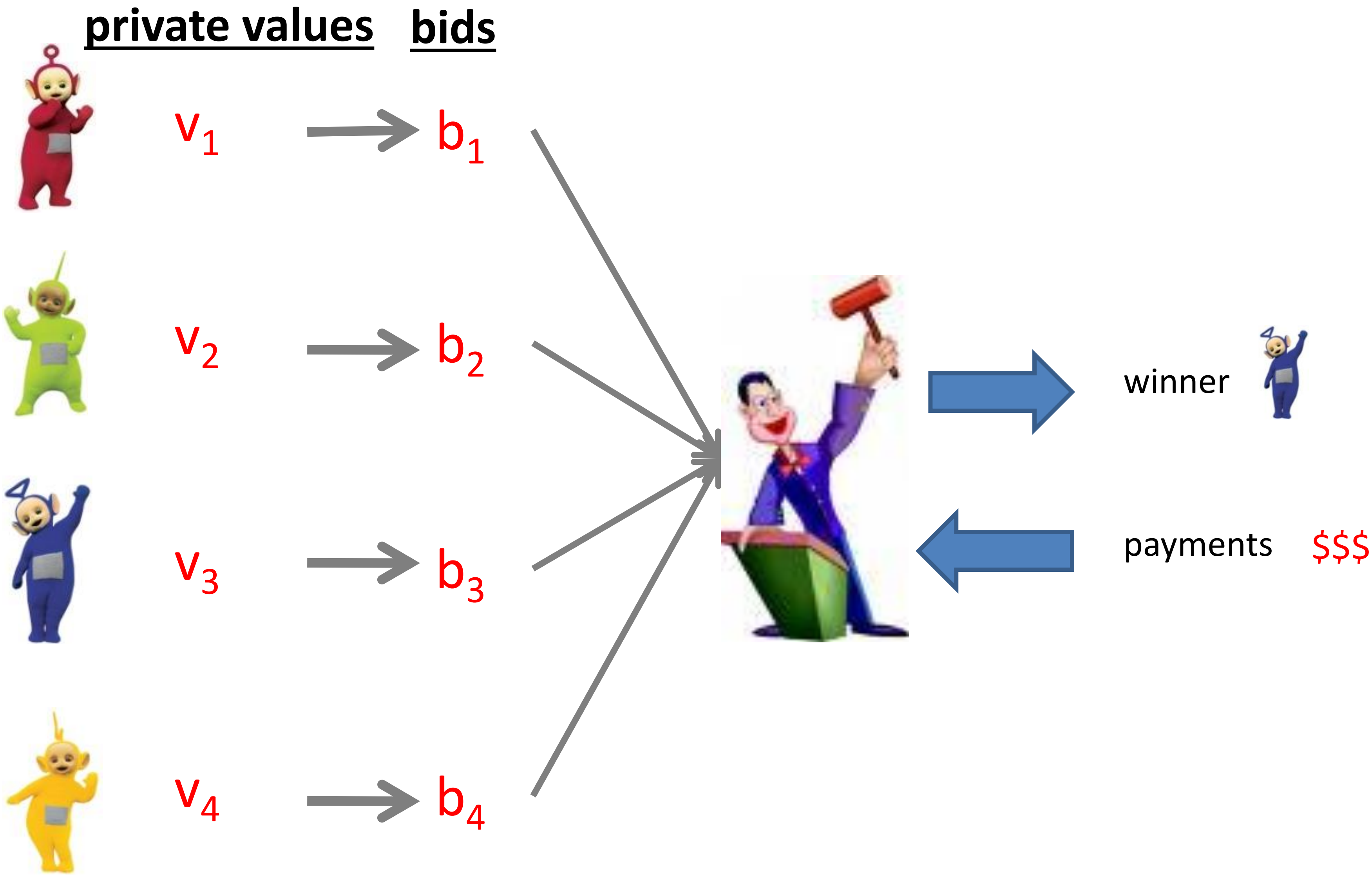
RTB Display Advertising Mechanism



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

Table of contents

Auctions scheme



Modeling

- n bidders

- Each bidder *i* 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.

- <u>Note:</u> 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, $b_i(v_i) = v_i+17$

	B(v)=v	**B(v)=v /2**	**B(v)=v /n**	**….**
B(v)=v				
…				

- Can be modeled as normal form game, where these strategies are the pure strategies.
- Example for a *game with incomplete information*.

Strategies and equilibrium

- An equilibrium in the auction is a profile of strategies $B_1, B_2, \ldots, B_n$ 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

- Truthful($b_i = v_i$)? NO!

$31

$30 ~~$100~~

Equilibrium in 1st-price auctions

- Suppose bidder *i*'s value is v_i in [0,1], which is only known by bidder *i*.
- Given this value, bidder *i* must submit a sealed bid $b_i(v_i)$
- We view bidder *i*'s strategy as a bidding function b_i : [0,1] -> $R_{+.}$ Some properties:
  - Bidders with higher values will place higher bids. So b_i is a strictly increasing function
  - Bidders are also *symmetric*. So bidders with the same value will submit the same bid: $b_i = b$ (*symmetric Nash equilibrium*)
  - Win(b_i) = F(v_i), where F is the C.D.F. of the true value distribution

Equilibrium in 1st-price auctions

- Bidder *1*'s payoff

$$\begin{cases} v_1 - b_1 & if\ \ b_1 > \max\{b(v_2),\ldots,b(v_n)\} \\ 0 & if\ \ b_1 \le \max\{b(v_2),\ldots,b(v_n)\} \end{cases}$$

- The expected payoff of bidding b_1 is given by

$$\begin{aligned} p(b_1) &= (v_1 - b_1)P(b_1 > \max\{b(v_2),\ldots,b(v_n)\}) \\ &= (v_1 - b_1)P(b_1 > b(v_2),\ldots,b_1 > (v_n)) \end{aligned}$$

- An optimal strategy b_i should maximize $p(b_1)$

Equilibrium in 1st-price auctions

- Suppose that bidder *i* cannot attend the auction and that she asks a friend to bid for her
  - The friend knows the equilibrium bidding function b^* but doe not know v_i
  - Bidder tells his friend the value as x and wants him to submit the bid $b^*(x)$
  - The expected pay off in this case is

$$p(b^*, x) = (v_1 - b^*(x))P(b^*(x) > b^*(v_2), \ldots, b^*(x) > b^*(v_n))$$
$$= (v_1 - b^*(x))P(x > v_2, \ldots, x > v_n) = (v_1 - b^*(x))F^{N-1}(x)$$

- The expected payoff is maximized when reporting his true value v_i to his friend ($x = v_i$)

Equilibrium in 1st-price auctions

- So if we differentiate the expected payoff with respect to x, the resulting derivative must be zero when $x = v_i$:

$$\frac{dp(b^*, x)}{dx} = \frac{d(v_1 - b^*(x))F^{N-1}(x)}{dx}$$

$$= (N-1)F^{N-2}(x)f(x)(v_1 - b^*(x)) - F^{N-1}(x)b^{*}{'}(x)$$

- The above equals zero when $x = v_i$; rearranging yields: $(N-1)F^{N-2}(v_1)f(v_1)v_1$

$$= F^{N-1}(v_1)b^{*}{'}(v_1) + (N-1)F^{N-2}(v_1)f(v_1)b^*(v_1)$$

$$= \frac{dF^{N-1}(v_1)b^*(v_1)}{dv}$$

Equilibrium in 1st-price auctions

- Taking the integration on both side

$$F^{N-1}(v_1)b^*(v_1) = (N-1)\int_0^{v_i} xf(x)F^{N-2}(x)\,dx + \text{constant}$$

- If we assume a bidder with value zero must bid zero, the above constant is zero. Therefore, we have (replace v_i with v)

$$b^*(v) = \frac{(N-1)\int_0^{v} xf(x)F^{N-2}(x)\,dx}{F^{N-1}(v)} = \frac{\int_0^{v} x\,dF^{N-1}(x)}{F^{N-1}(v)}$$

- It shows that in the equilibrium, each bidder bids the expectation of the second-highest bidder's value conditional on winning the auction.

Untruthful bidding in 1st-price auctions

- Suppose that each bidder's value is uniformly distributed on [0,1].
  - Replacing F(v)=v and f(v)=1 gives

$$b^*(v) = \frac{\int_0^v x\,dF^{N-1}(x)}{F^{N-1}(v_1)} = \frac{\int_0^v x\,dx^{N-1}}{v^{N-1}}$$

$$= \frac{\int_0^v x(N-1)x^{N-2}\,dx}{v^{N-1}} = \frac{(N-1)\int_0^v x^{N-1}\,dx}{v^{N-1}}$$

$$= \frac{(N-1)}{v^{N-1}}\frac{1}{N}v^N = v - \frac{v}{N}$$

Equilibrium in 2nd-price auctions

- bidder *1*'s payoff

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

- 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 $b_1 > v_1$

Equilibrium in 2nd-price auctions

- bidder *1*'s payoff

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

- The expected payoff of bidding b_1 is given by

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

- Or taking derivative of $\pi(v_1, b_1)$ w.r.t. b1 yields $b_1 = v_1$

So telling the truth $b_1 = v_1$ is a Bayesian Nash equilibrium bidding strategy!

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

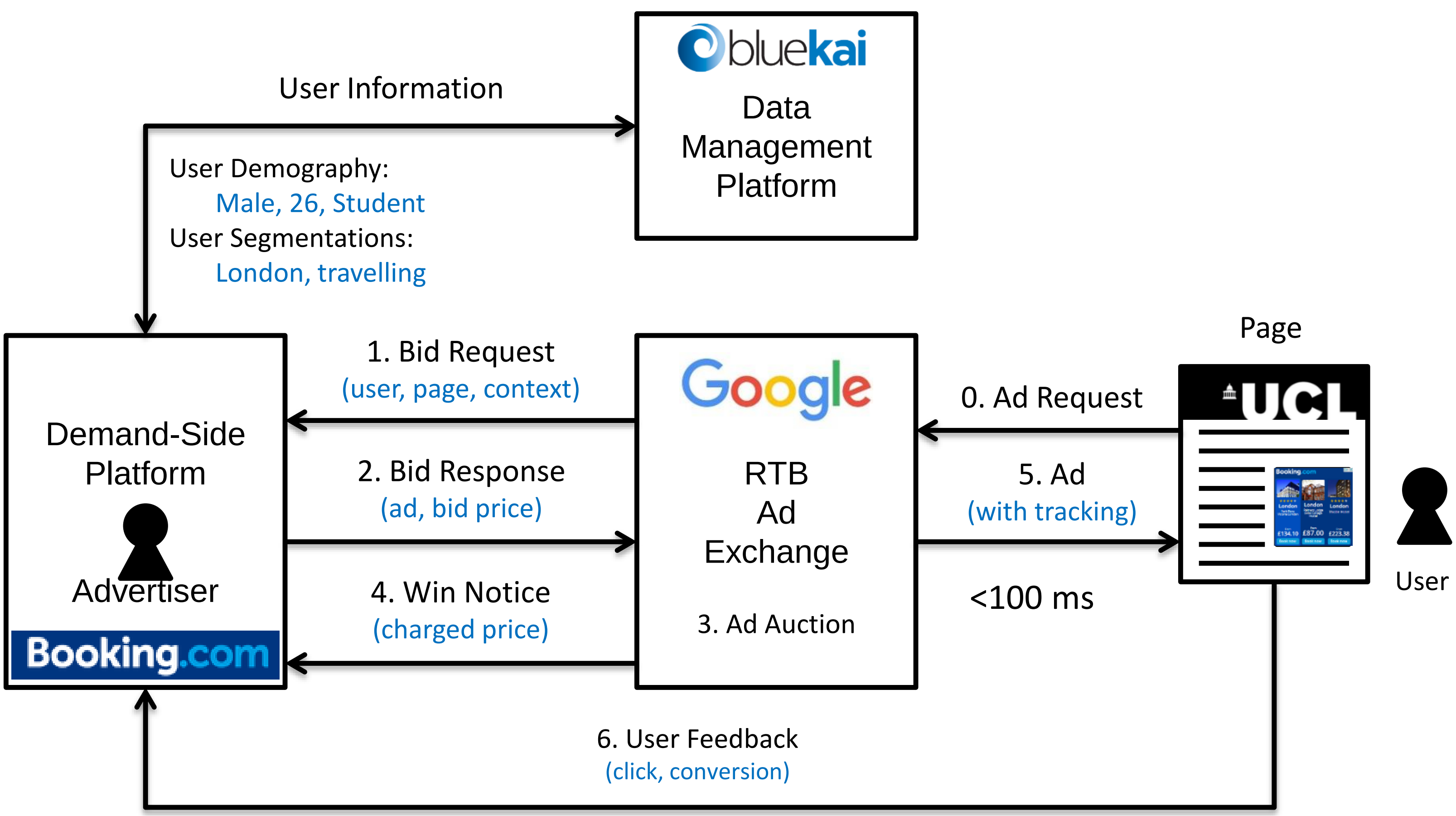
RTB Auctions

- Second price auction with reserve price
- From a bidder's perspective, the market price z refers to the highest bid from competitors
- Payoff: $(v_{impression} - z) \times P(win)$
- Value of impression depends on user response

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Learning to bid
- Conversion attribution
- Pacing control
- Targeting and audience expansion
- Reserve price optimization

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.

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.

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

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.

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=[Weekday=Wednesday, Gender=Male, City=London]

x=[0,0,1,0,0,0,0 0,1 0,0,1,0…0]

High dimensional sparse binary feature vector

Linear Models

- Logistic Regression
  - With SGD learning
  - Sparse solution

- Online Bayesian Probit Regression

ML Framework of CTR Estimation

- A binary regression problem

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

  - 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
      - An isotonic mapping from prediction to calibrated prediction

Logistic Regression

- Prediction

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

- Cross Entropy Loss

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

- Stochastic Gradient Descent Learning

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

[Lee et al. Estimating Conversion Rate in Display Advertising from Past Performance Data. KDD 12]

Logistic Regression with SGD

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

- 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

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

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

adaptively selects regularisation functions

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

t: current example index

g_s: gradient for example t

- Online closed-form update of FTRL

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

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

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

[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]

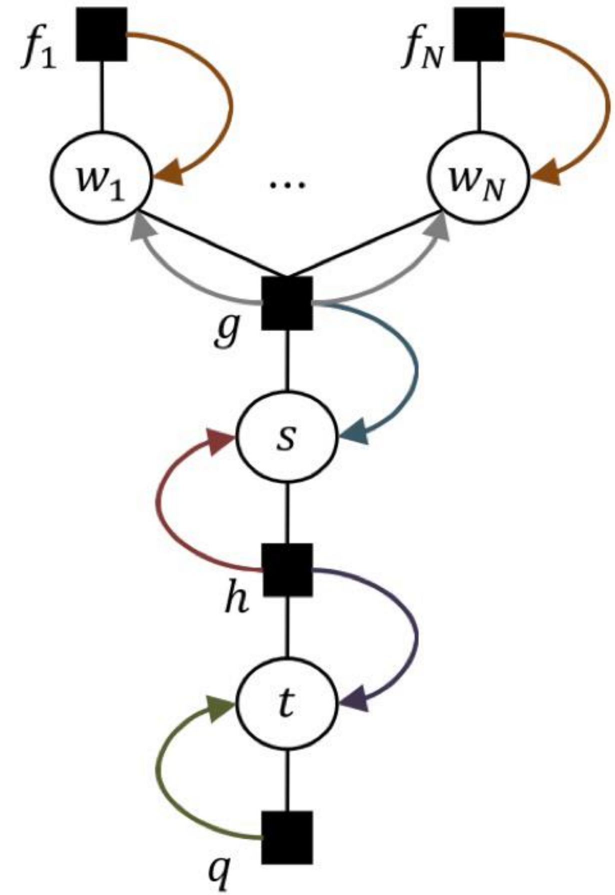
Online Bayesian Probit Regression

Given feature x, predicting click y

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

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

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



Approximated inference via Expectation Propagation

The factorised model

$$p(y \mid t) \cdot p(t \mid s) \cdot p(s \mid \boldsymbol{x}, \boldsymbol{w}) \cdot p(\boldsymbol{w})$$

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

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

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

[Graepel et al. Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine. ICML 10]

Linear Prediction Models

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

- 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

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

Factorisation Machines

- Prediction based on feature embedding

$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\Big(\underbrace{w_0 + \sum_{i=1}^{N} w_i x_i}_{\text{Logistic Regression}} + \underbrace{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j}_{\text{Feature Interactions}}\Big)$$

  - 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]

Factorisation Machines

- Prediction based on feature embedding

$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\Big(\underbrace{w_0 + \sum_{i=1}^{N} w_i x_i}_{\text{Logistic Regression}} + \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j}_{\text{Feature Interactions}}\Big)$$

For x=[Weekday=Friday, Gender=Male, City=Shanghai]

$$y_{\mathrm{FM}}(\boldsymbol{x}) = \mathrm{sigmoid}\Big(w_0 + w_{\mathrm{Friday}} + w_{\mathrm{Male}} + w_{\mathrm{Shanghai}} + \langle \boldsymbol{v}_{\mathrm{Friday}}, \boldsymbol{v}_{\mathrm{Male}} \rangle + \langle \boldsymbol{v}_{\mathrm{Friday}}, \boldsymbol{v}_{\mathrm{Shanghai}} \rangle + \langle \boldsymbol{v}_{\mathrm{Male}}, \boldsymbol{v}_{\mathrm{Shanghai}} \rangle\Big)$$

[Rendle. Factorization machines. ICDM 2010.]

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

Field-aware Factorisation Machines

- Feature embedding for another field

$$y_{\mathrm{FFM}}(\boldsymbol{x}) = \mathrm{sigmoid}\Big(w_0 + \sum_{i=1}^{N} w_i + \boxed{\sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle \boldsymbol{v}_{i,\mathrm{field}(j)}, \boldsymbol{v}_{j,\mathrm{field(i)}} \rangle x_i x_j}\Big)$$

Field-aware field embedding

For x=[Weekday=Friday, Gender=Male, City=Shanghai]

$$\begin{aligned} y_{\mathrm{FFM}}(\boldsymbol{x}) = \mathrm{sigmoid}\Big(& w_0 + w_{\mathrm{Friday}} + w_{\mathrm{Male}} + w_{\mathrm{Shanghai}} \\ & + \langle \boldsymbol{v}_{\mathrm{Friday,Gender}}, \boldsymbol{v}_{\mathrm{Male,Weekday}} \rangle + \langle \boldsymbol{v}_{\mathrm{Friday,City}}, \boldsymbol{v}_{\mathrm{Shanghai,Weekday}} \rangle \\ & + \langle \boldsymbol{v}_{\mathrm{Male,City}}, \boldsymbol{v}_{\mathrm{Shanghai,Gender}} \rangle \Big) \end{aligned}$$

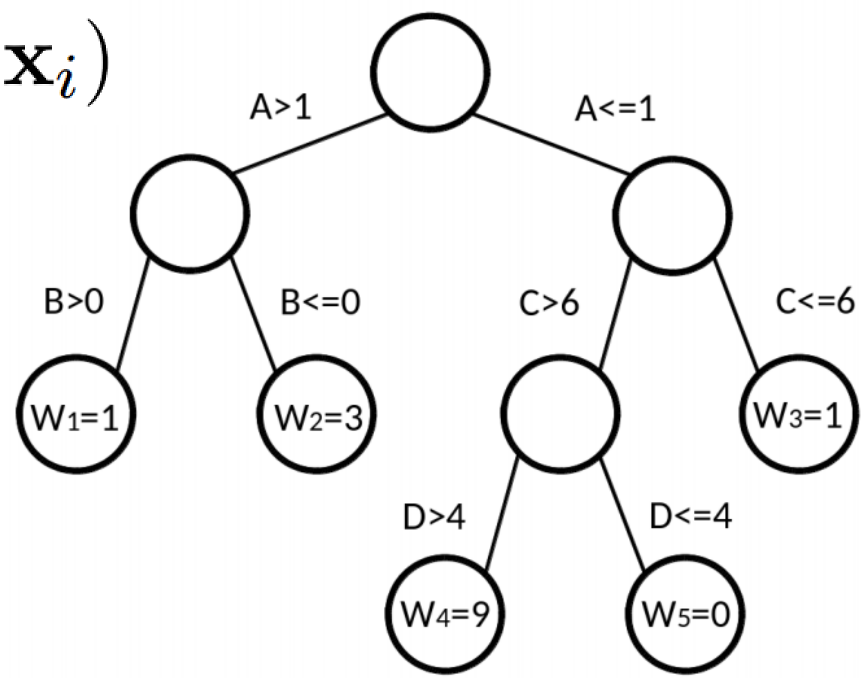
[Juan et al. Field-aware Factorization Machines for CTR Prediction. RecSys 2016.]

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 $f_k(\mathbf{x}_i)$



[Chen and He. Higgs Boson Discovery with Boosted Trees . HEPML 2014.]

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

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

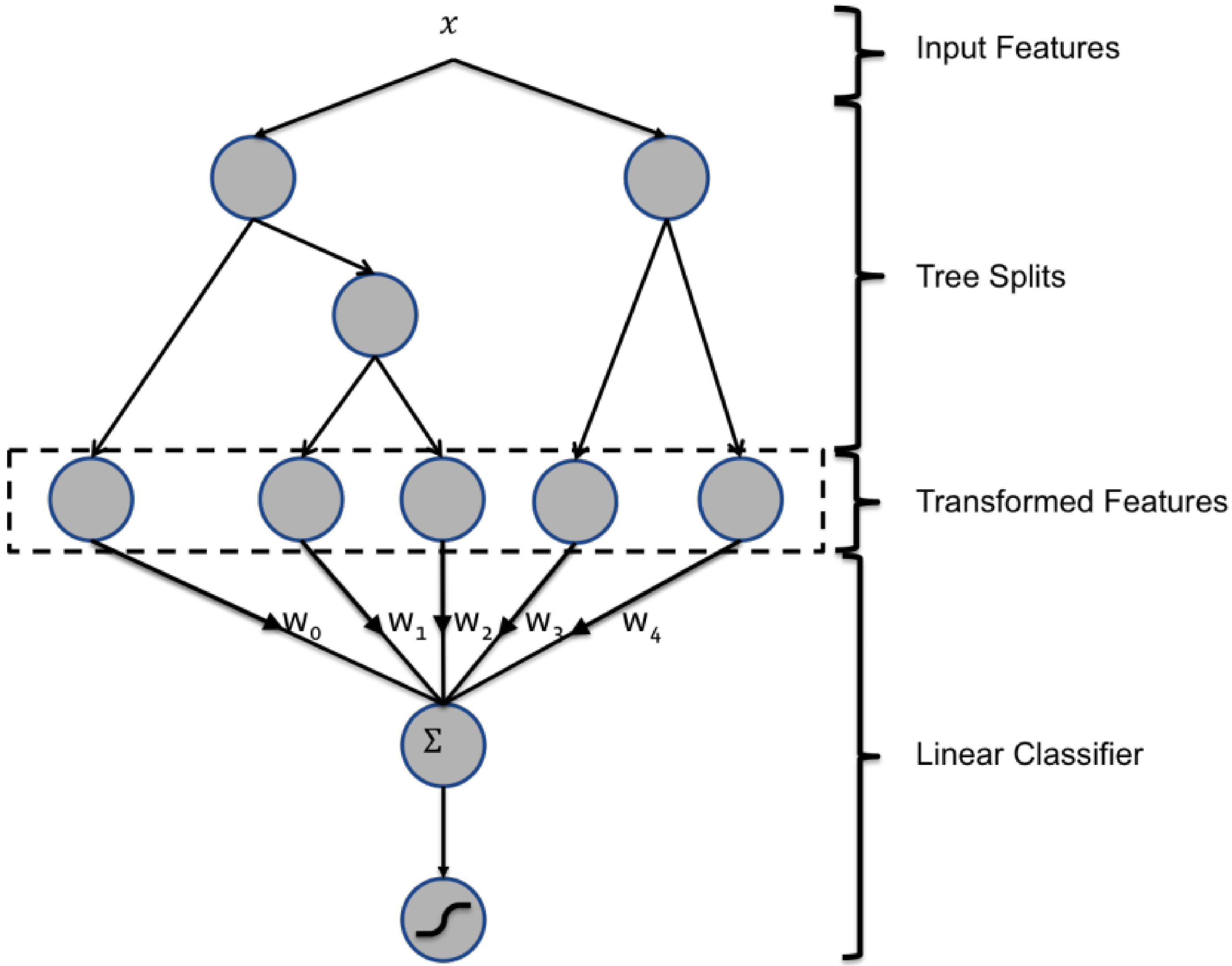
$$= \sum_{i=1}^{n} l(y_i, \hat{y_i}^{(t-1)} + f_t(\mathbf{x}_i)) + \sum_{i=1}^{t} \Omega(f_i)$$

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

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

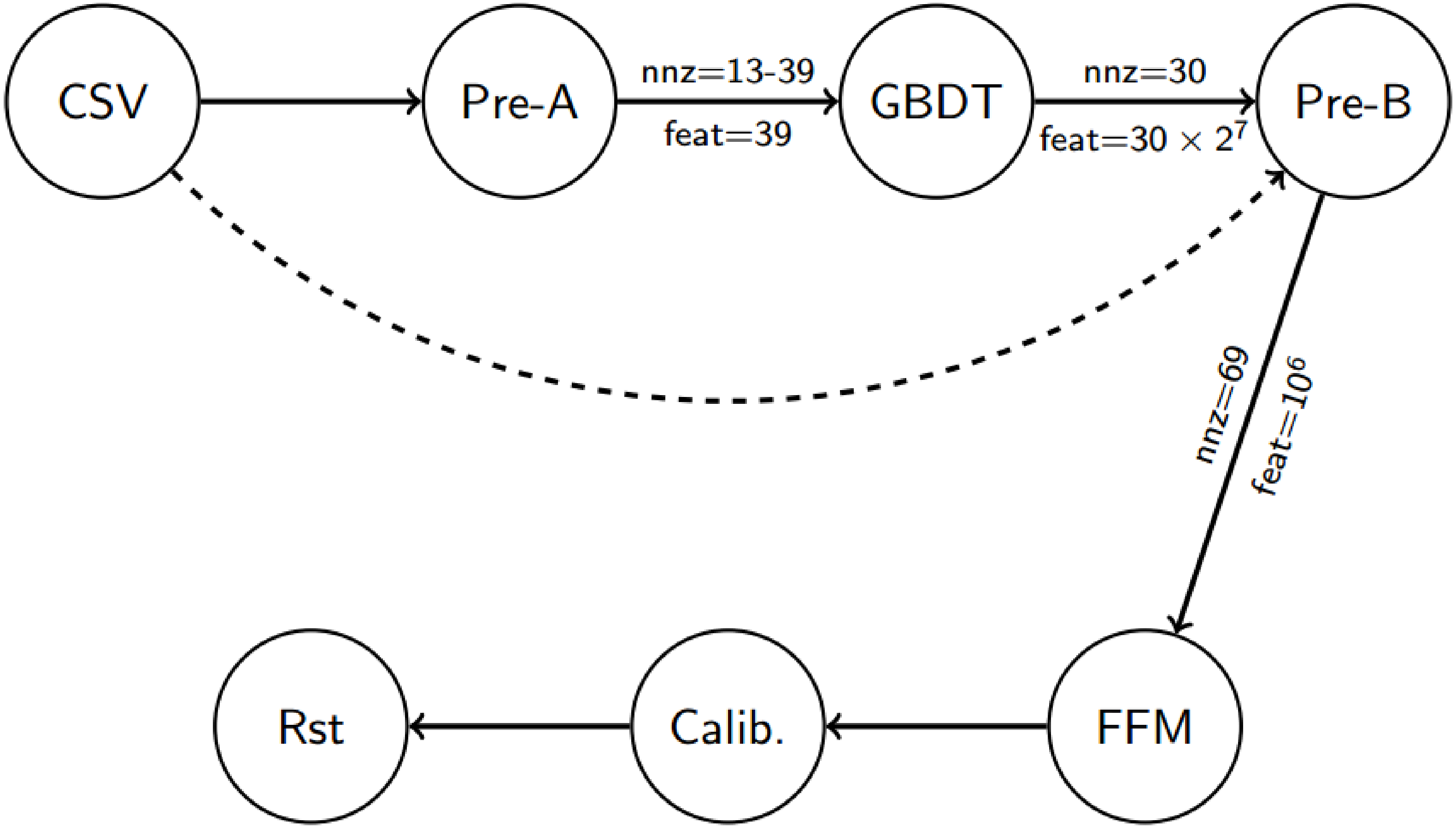
[Tianqi Chen. https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf]
[Chen and He. Higgs Boson Discovery with Boosted Trees . HEPML 2014.]

Combined Models: GBDT + LR



[He et al. Practical Lessons from Predicting Clicks on Ads at Facebook . ADKDD 2014.]

Combined Models: GBDT + FM

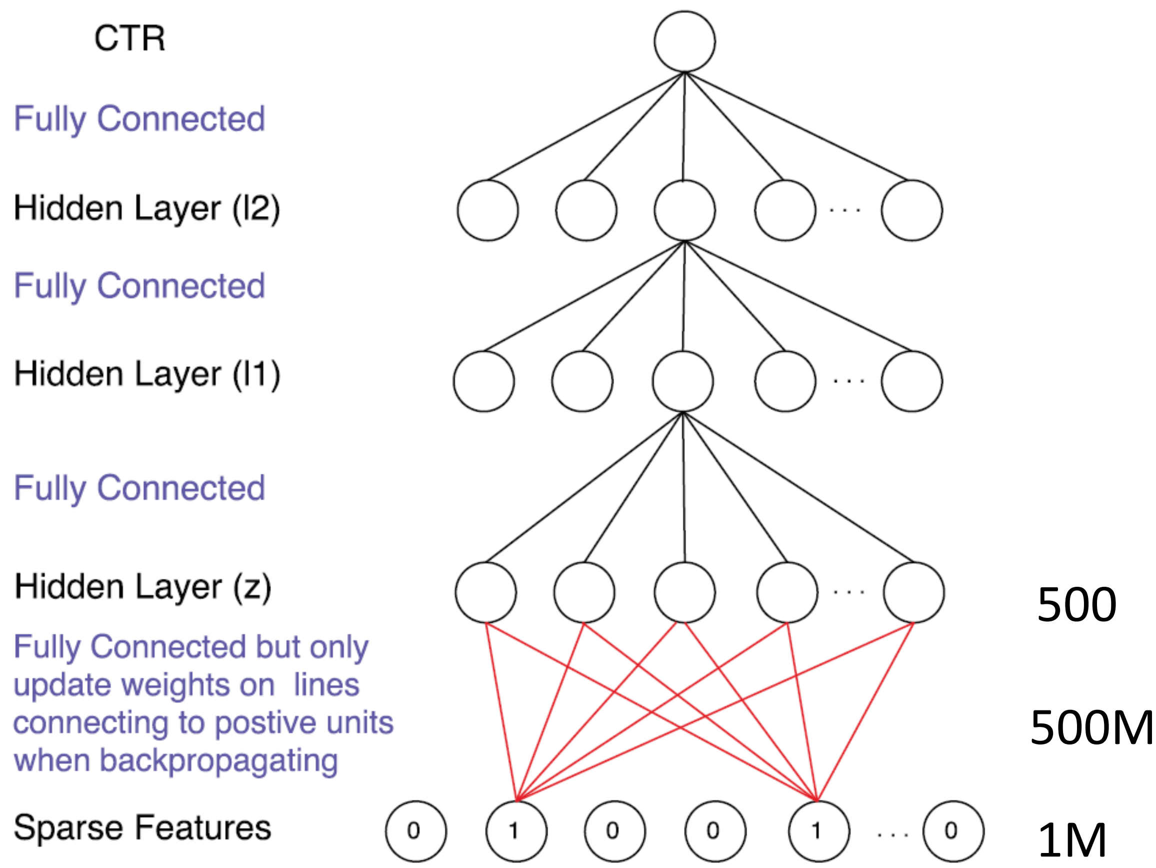


"nnz" means the number of non-zero elements of each impression; "feat" represents the size of feature space.

[http://www.csie.ntu.edu.tw/~r01922136/kaggle-2014-criteo.pdf]

Neural Network Models

- Difficulty: Impossible to directly deploy neural network models on such data



E.g., input features 1M, first layer 500, then 500M parameters for first layer

Review Factorisation Machines

- Prediction based on feature embedding

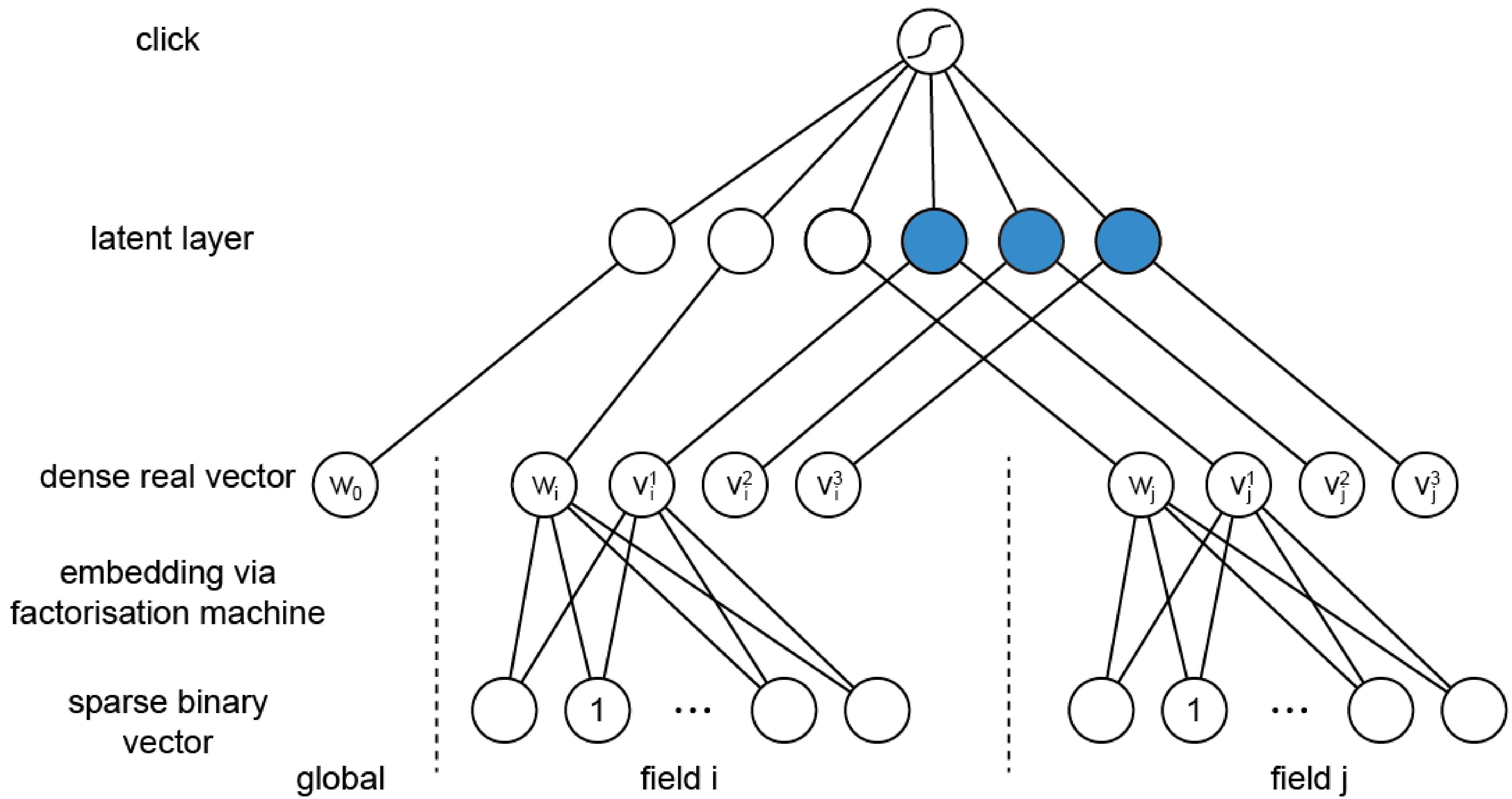
$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\Big(\underbrace{w_0 + \sum_{i=1}^{N} w_i x_i}_{\text{Logistic Regression}} + \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j}_{\text{Feature Interactions}}\Big)$$

  - Embed features into a k-dimensional latent space
  - Explore the feature interaction patterns using vector inner-product

[Rendle. Factorization machines. ICDM 2010.]

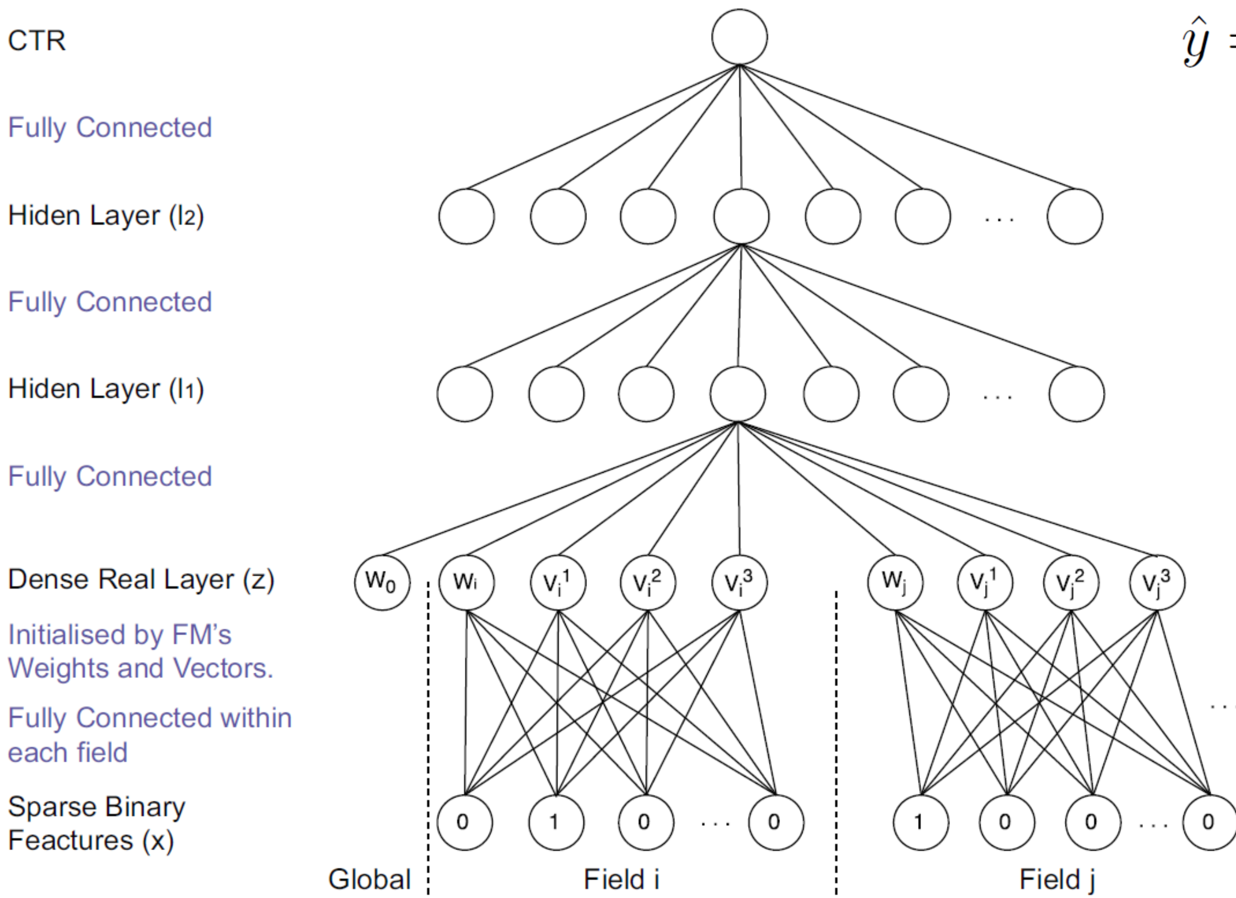
[Oentaryo et al. Predicting response in mobile advertising with hierarchical importance-aware factorization machine. WSDM 14]

Factorisation Machine is a Neural Network



$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\left(w_0 + \sum_{i=1}^{N} w_i x_i + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j\right)$$

Factorisation-machine supported Neural Networks (FNN)



$$\hat{y} = \mathrm{sigmoid}(\boldsymbol{W}_3 \boldsymbol{l}_2 + b_3)$$

$$\boldsymbol{l}_2 = \tanh(\boldsymbol{W}_2 \boldsymbol{l}_1 + \boldsymbol{b}_2)$$

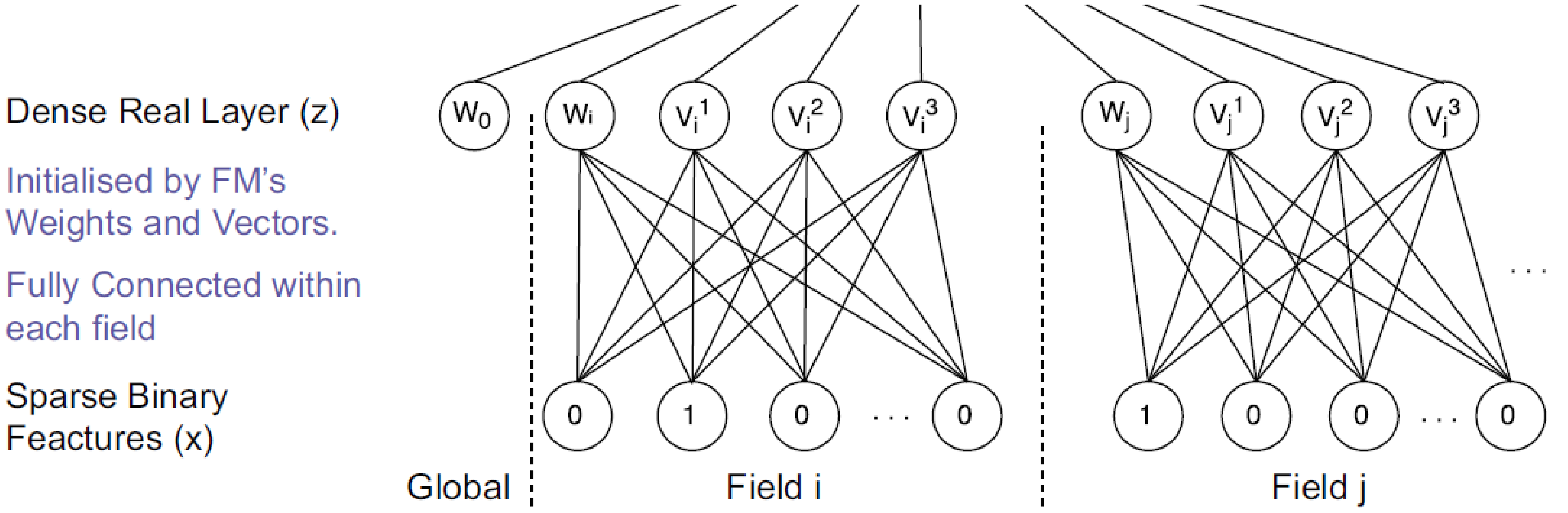
$$\boldsymbol{l}_1 = \tanh(\boldsymbol{W}_1 \boldsymbol{z} + \boldsymbol{b}_1)$$

$$\boldsymbol{z}_i = (w_i, v_i^1, v_i^2, \ldots, v_i^K) = \boldsymbol{W}_0^i \cdot \boldsymbol{x}[\mathrm{start}_i : \mathrm{end}_i]$$

[Factorisation Machine Initialised]

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

Factorisation-machine supported Neural Networks (FNN)



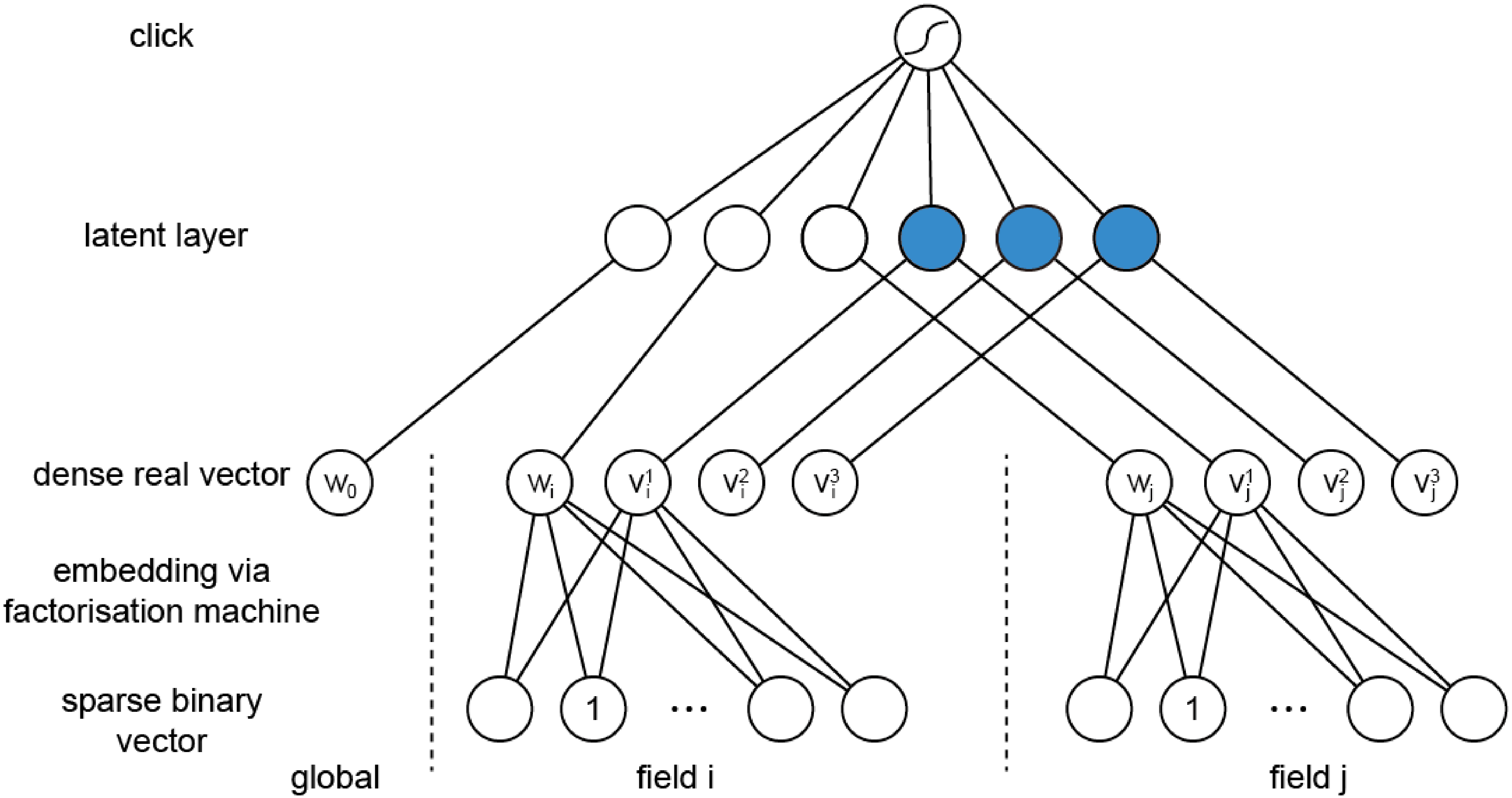
- Chain rule to update factorisation machine parameters

$$\frac{\partial L(y,\hat{y})}{\partial \boldsymbol{W}_0^i} = \frac{\partial L(y,\hat{y})}{\partial \boldsymbol{z}_i} \frac{\partial \boldsymbol{z}_i}{\partial \boldsymbol{W}_0^i} = \frac{\partial L(y,\hat{y})}{\partial \boldsymbol{z}_i} \boldsymbol{x}[\mathrm{start}_i : \mathrm{end}_i]$$

$$\boldsymbol{W}_0^i \leftarrow \boldsymbol{W}_0^i - \eta \cdot \frac{\partial L(y,\hat{y})}{\partial \boldsymbol{z}_i} \boldsymbol{x}[\mathrm{start}_i : \mathrm{end}_i].$$

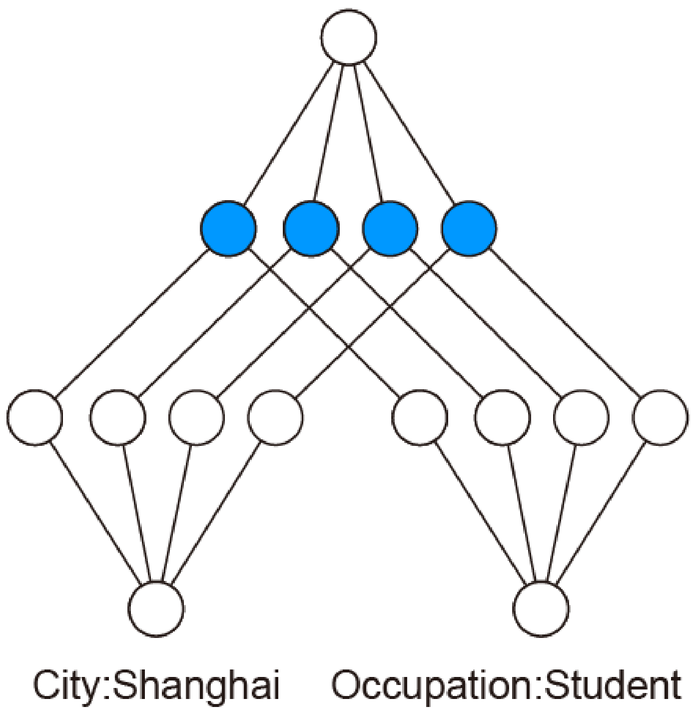
[Zhang et al. Deep Learning over Multi-field Categorical Data – A Case Study on User Response Prediction. ECIR 16]

But factorisation machine is still different from common additive neural networks

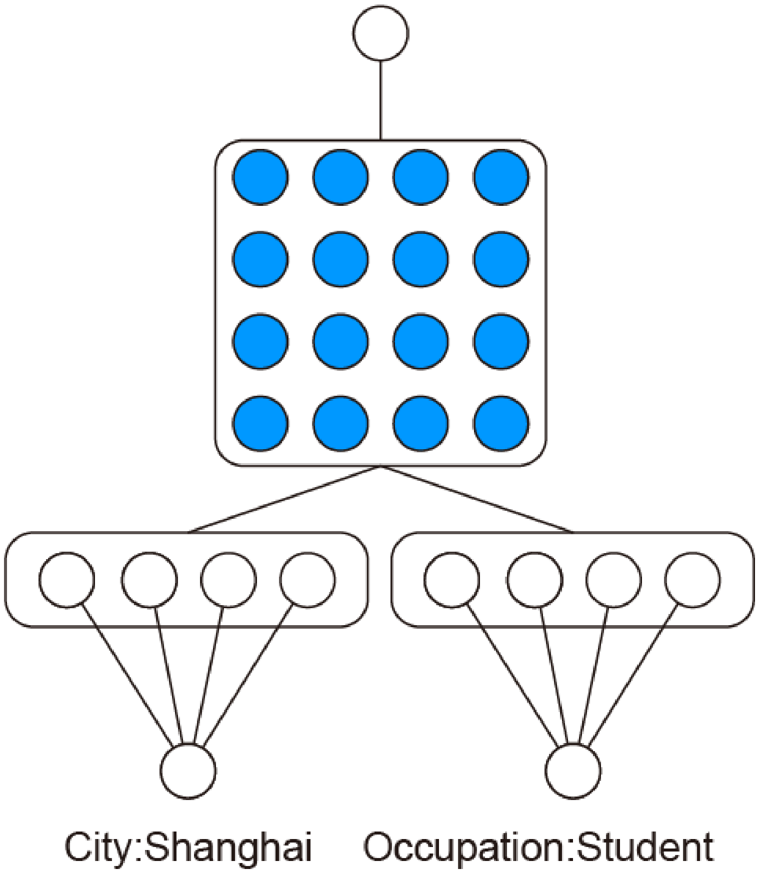


$$y_{\mathrm{FM}}(\boldsymbol{x}) := \mathrm{sigmoid}\Big(w_0 + \sum_{i=1}^{N} w_i x_i + \sum_{i=1}^{N}\sum_{j=i+1}^{N} \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle x_i x_j\Big)$$

Product Operations as Feature Interactions



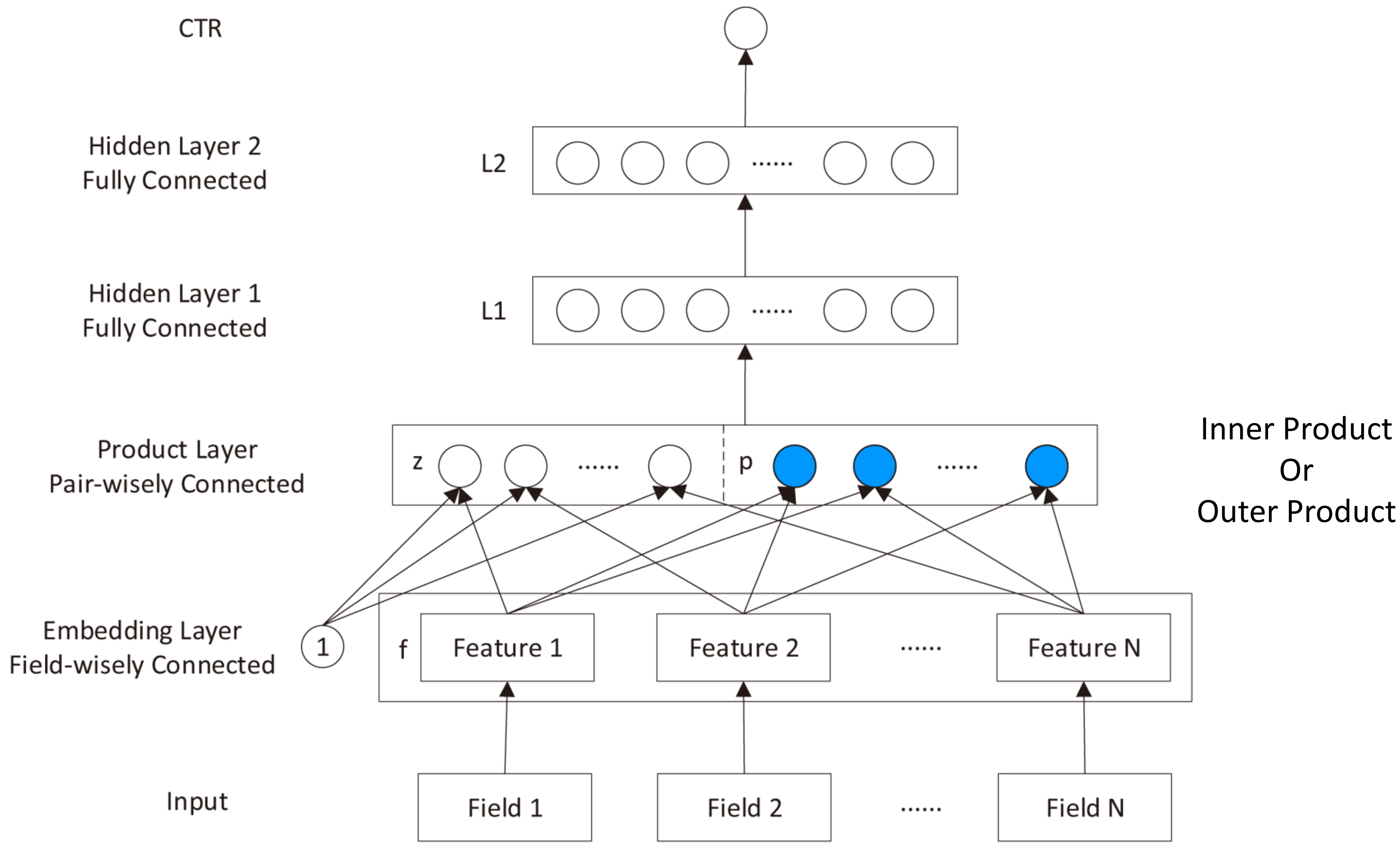
Inner Product Operation



Outer Product Operation

[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. ICDM 2016]

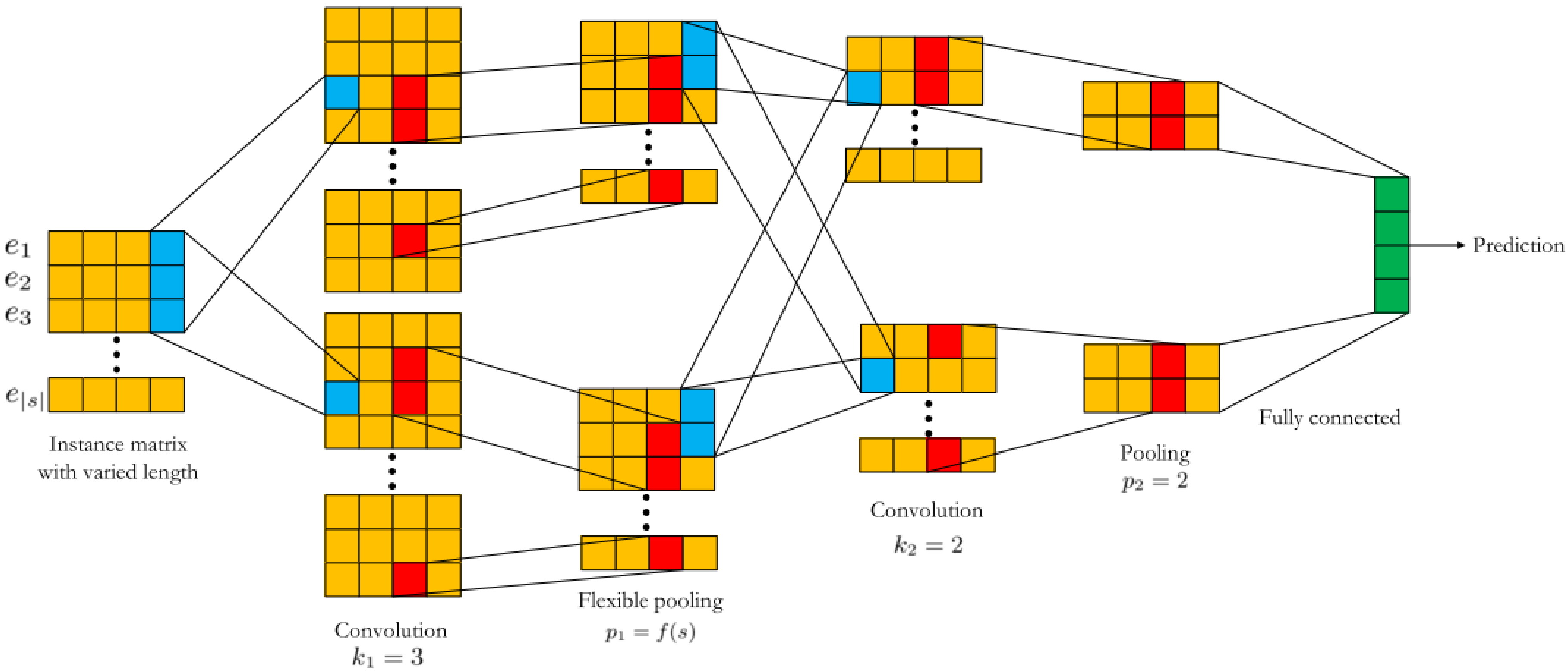
Product-based Neural Networks (PNN)



[Yanru Qu et al. Product-based Neural Networks for User Response Prediction. ICDM 2016]

Convolutional Click Prediction Model (CCPM)

- CNN to (partially) select good feature combinations



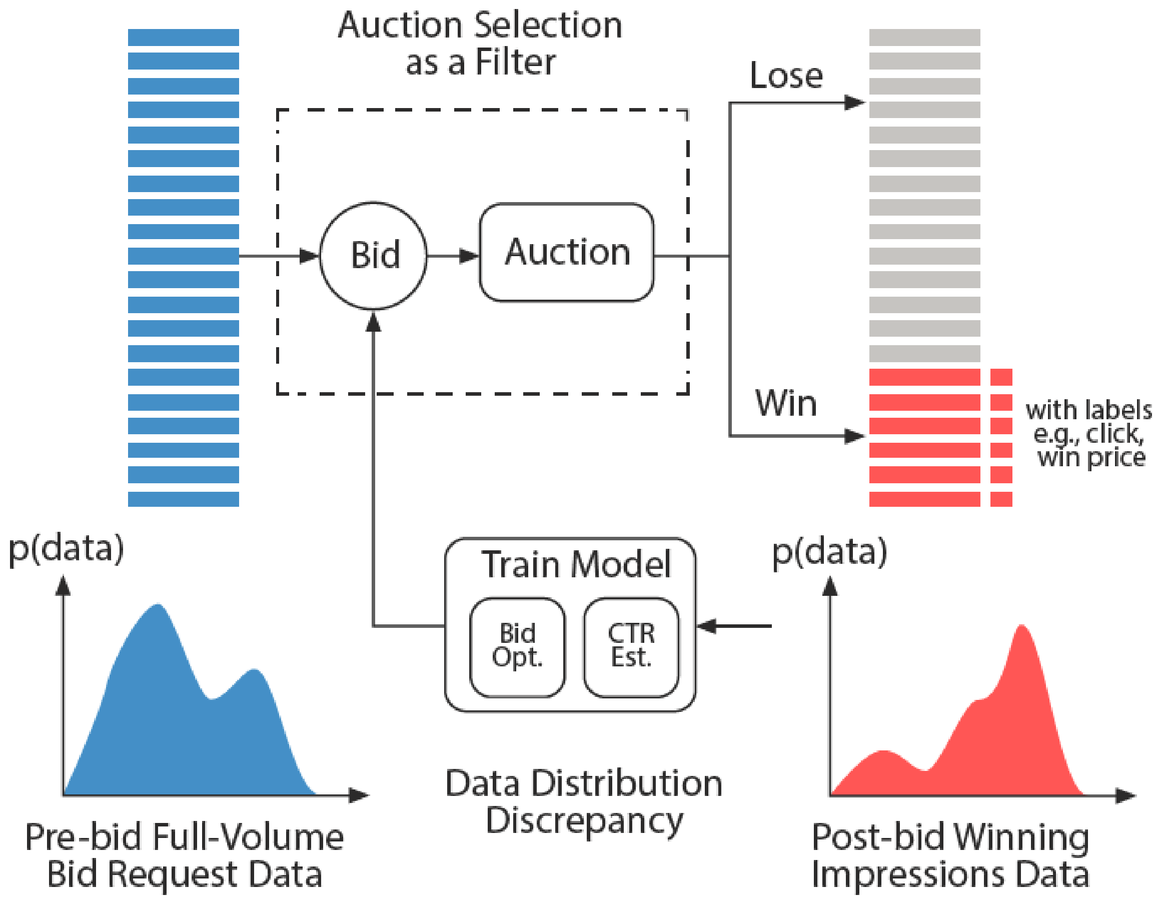
[Qiang Liu et al. A convolutional click prediction model. CIKM 2015]

Overall Performance

Model	AUC		Log Loss	
	Criteo	iPinYou	Criteo	iPinYou
LR	71.48%	73.43%	0.1334	5.581e-3
FM	72.20%	75.52%	0.1324	5.504e-3
FNN	75.66%	76.19%	0.1283	5.443e-3
CCPM	76.71%	76.38%	0.1269	5.522e-3
PNN-I	**77.79%**	79.14%	**0.1252**	5.195e-3
PNN-II	77.54%	**81.74%**	0.1257	5.211e-3
PNN-III	77.00%	76.61%	0.1270	**4.975e-3**

Model	RMSE		RIG	
	Criteo	iPinYou	Criteo	iPinYou
LR	9.362e-4	5.350e-07	6.680e-2	7.353e-2
FM	9.284e-4	5.343e-07	7.436e-2	8.635e-2
FNN	9.030e-4	5.285e-07	1.024e-1	9.635e-2
CCPM	8.938e-4	5.343e-07	1.124e-1	8.335e-2
PNN-I	**8.803e-4**	4.851e-07	**1.243e-1**	1.376e-1
PNN-II	8.846e-4	5.293e-07	1.211e-1	1.349e-1
PNN-III	8.988e-4	**4.819e-07**	1.118e-1	**1.740e-1**

Training with Instance Bias



$$\underbrace{p_x(\boldsymbol{x})}_{\text{bid request}} \cdot \underbrace{P(\text{win}|\boldsymbol{x}, b_{\boldsymbol{x}})}_{\text{auction selection}} = \underbrace{q_x(\boldsymbol{x})}_{\text{impression}}$$

$$w(b_{\boldsymbol{x}}) \equiv P(\text{win}|\boldsymbol{x}, b_{\boldsymbol{x}}) = \int_0^{b_{\boldsymbol{x}}} p_z^{\boldsymbol{x}}(z)dz$$

[Zhang et al. Bid-aware Gradient Descent for Unbiased Learning with Censored Data in Display Advertising. KDD 2016.]

Unbiased Learning

- General machine learning problem

$$\min_{\boldsymbol{\theta}} \quad \mathbb{E}_{\boldsymbol{x} \sim p_x(\boldsymbol{x})}[\mathcal{L}(y, f_{\boldsymbol{\theta}}(\boldsymbol{x}))] + \lambda \Phi(\boldsymbol{\theta})$$

- But the training data distribution is *q(x)*
  - A straightforward solution: importance sampling

$$\mathbb{E}_{\boldsymbol{x} \sim p_x(\boldsymbol{x})}[\mathcal{L}(y, f_{\boldsymbol{\theta}}(\boldsymbol{x}))] = \int_{\boldsymbol{x}} p_x(\boldsymbol{x}) \mathcal{L}(y, f_{\boldsymbol{\theta}}(\boldsymbol{x})) d\boldsymbol{x}$$

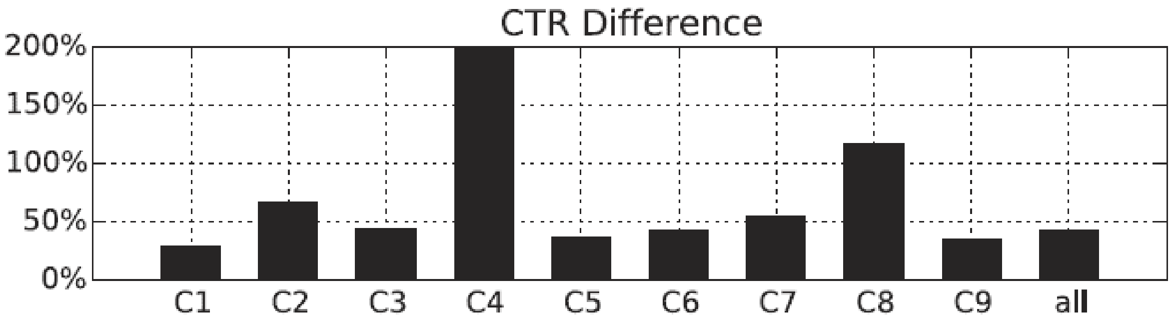
$$= \int_{\boldsymbol{x}} q_x(\boldsymbol{x}) \frac{\mathcal{L}(y, f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{w(b_{\boldsymbol{x}})} d\boldsymbol{x} = \mathbb{E}_{\boldsymbol{x} \sim q_x(\boldsymbol{x})}\Big[\frac{\mathcal{L}(y, f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{w(b_{\boldsymbol{x}})}\Big]$$

[Zhang et al. Bid-aware Gradient Descent for Unbiased Learning with Censored Data in Display Advertising. KDD 2016.]

Unbiased CTR Estimator Learning

Table : Online A/B testing of CTR estimation (Yahoo!).

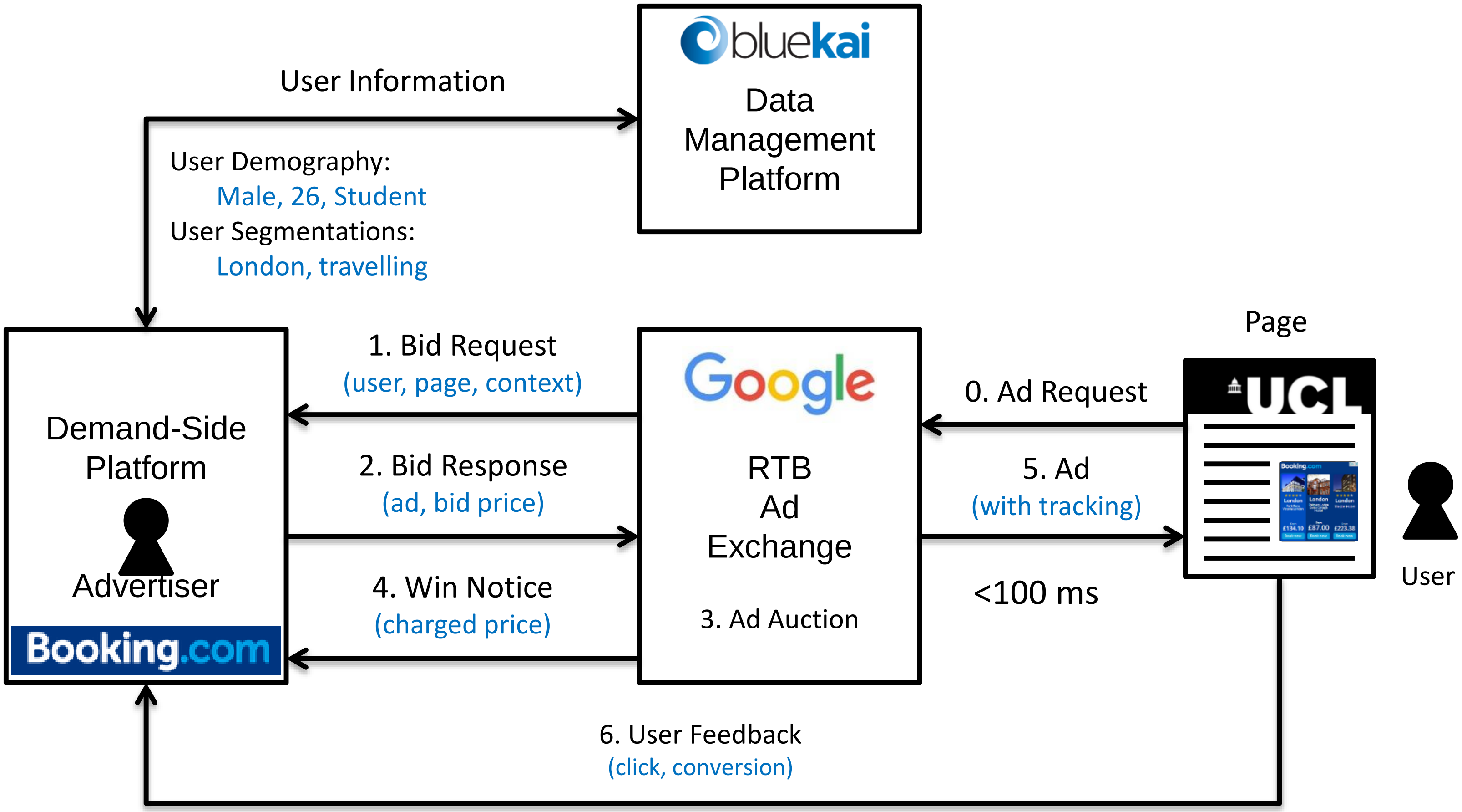
Camp.	BIAS AUC.	KMMP AUC	AUC Lift
C1	63.78%	64.12%	0.34%
C2	87.45%	88.58%	1.13%
C3	69.73%	75.52%	5.79%
C4	88.82%	89.55%	0.73%
C5	69.71%	72.29%	2.58%
C6	89.33%	90.70%	1.37%
C7	77.76%	78.92%	1.16%
C8	74.57%	76.98%	2.41%
C9	71.04%	73.12%	2.08%
all	73.48%	76.45%	2.97%



[Zhang et al. Bid-aware Gradient Descent for Unbiased Learning with Censored Data in Display Advertising. KDD 2016.]

Table of contents

RTB Display Advertising Mechanism



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

Data of Learning to Bid

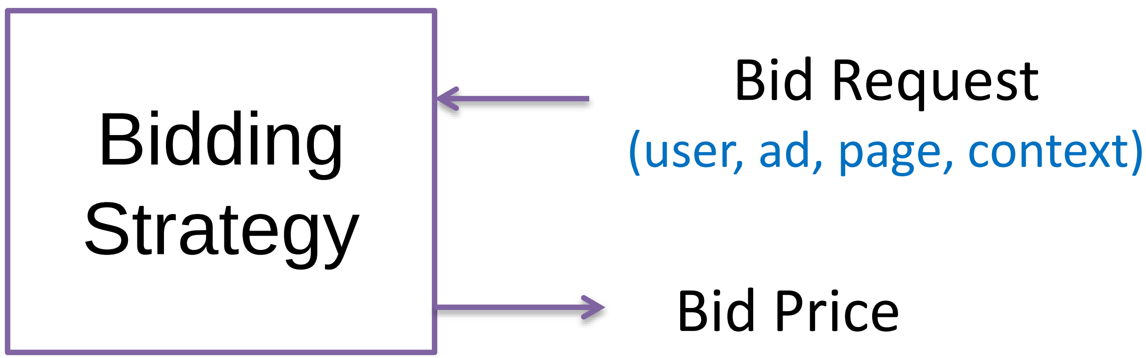
- Data

$(\mathbf{x}, t)$	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

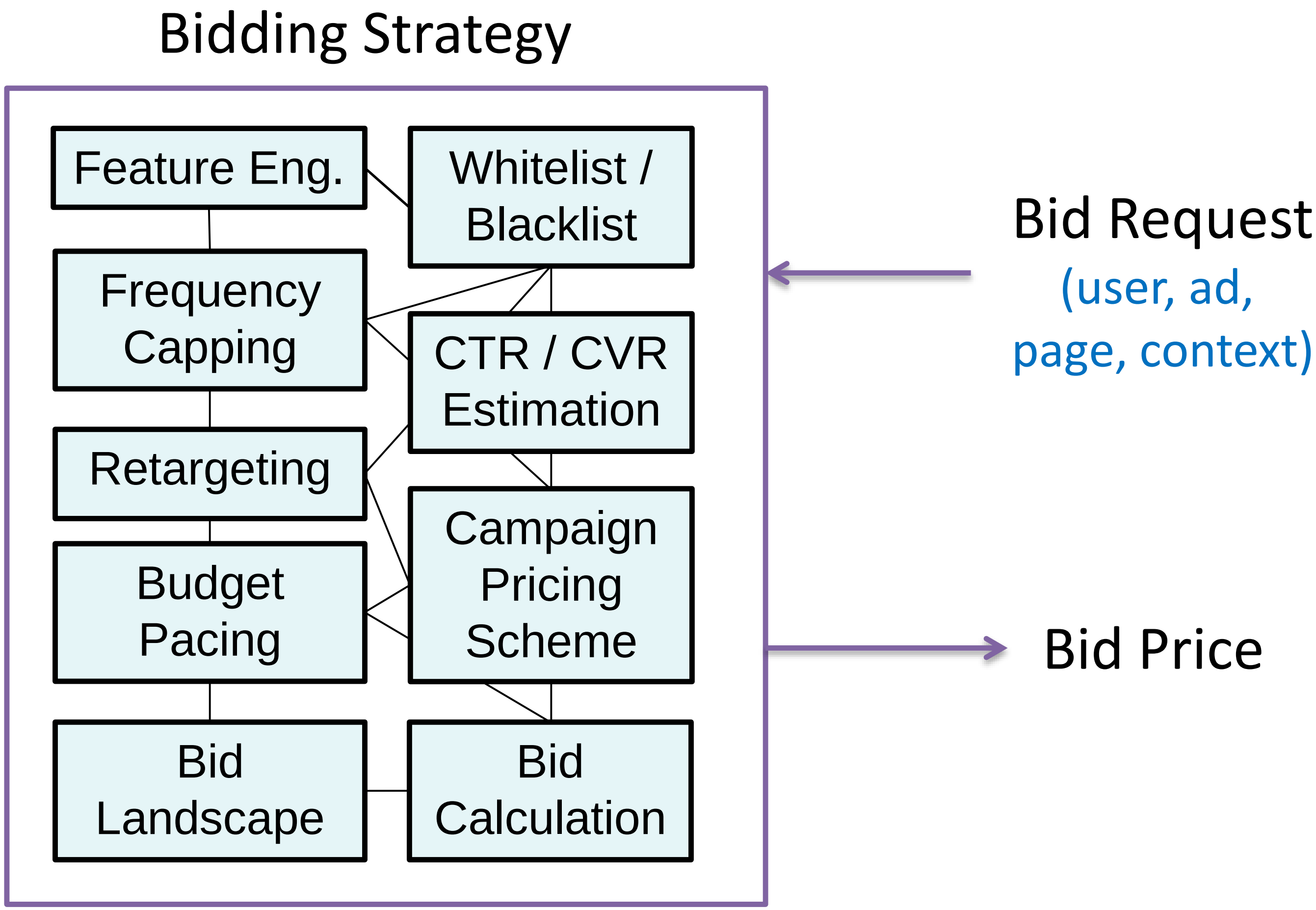
- How much to bid for each bid request?
  - Find an optimal bidding function b(x)



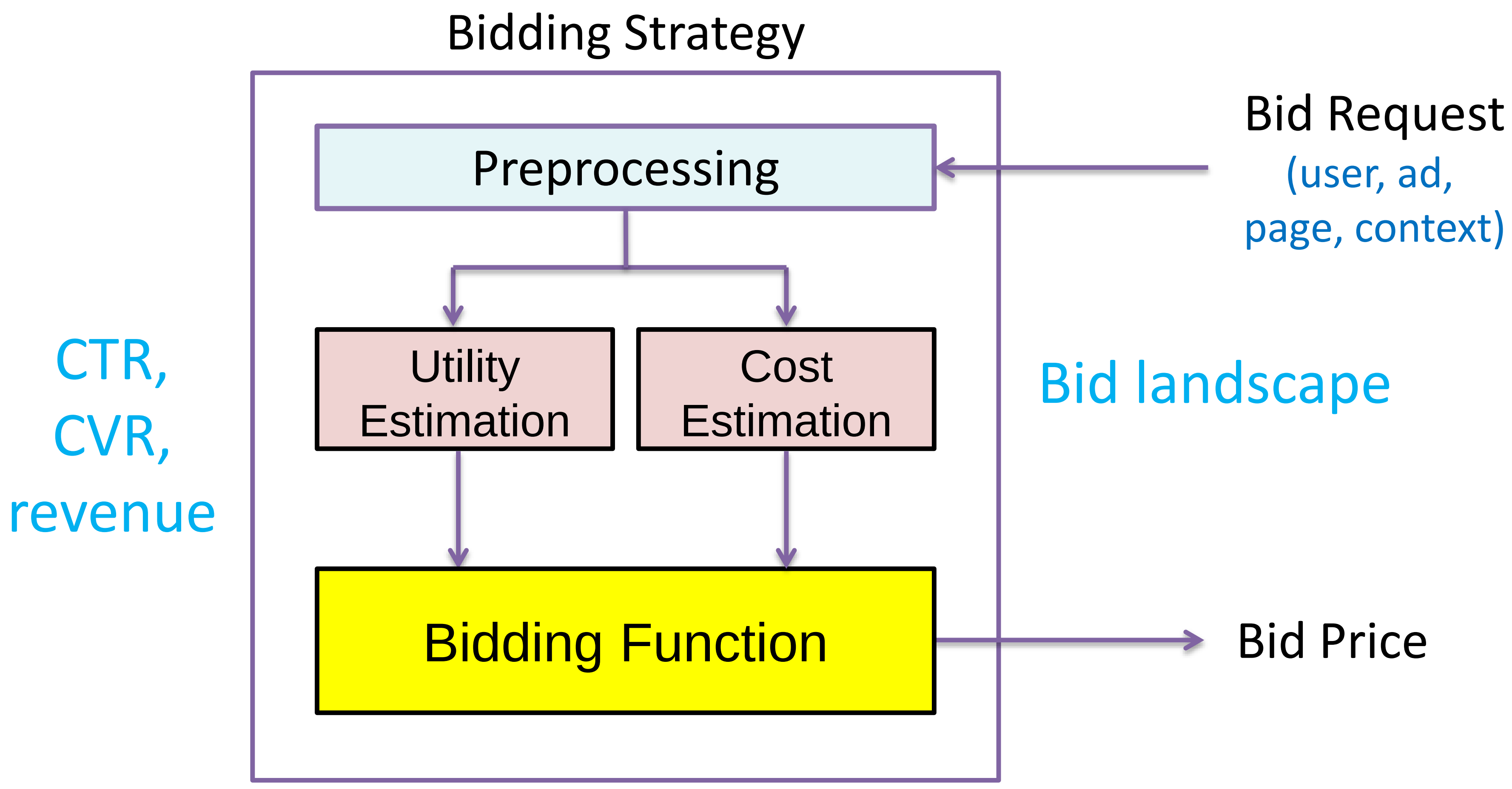
- Bid to optimise the KPI with budget constraint

$$\max_{\text{bidding strategy}} \quad \text{KPI}$$
$$\text{subject to} \quad \text{cost} \leq \text{budget}$$

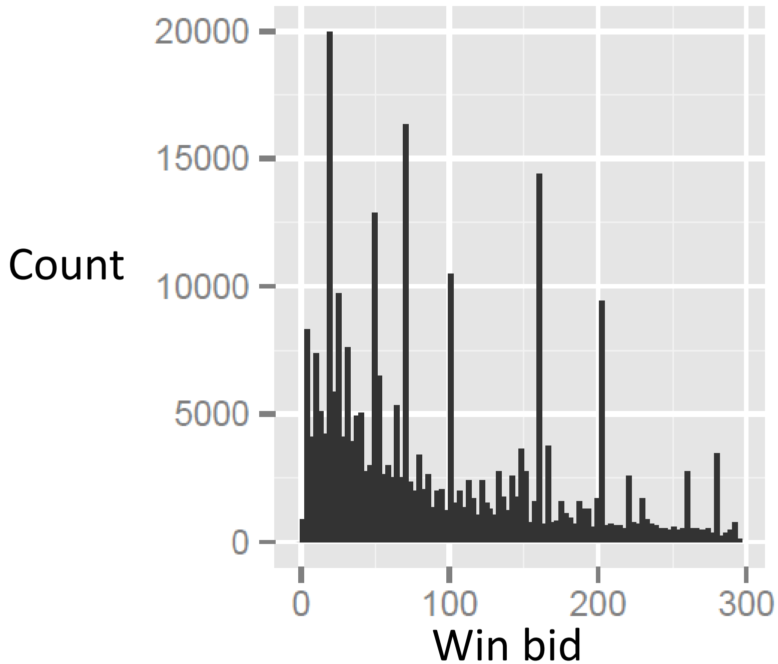
Bidding Strategy in Practice

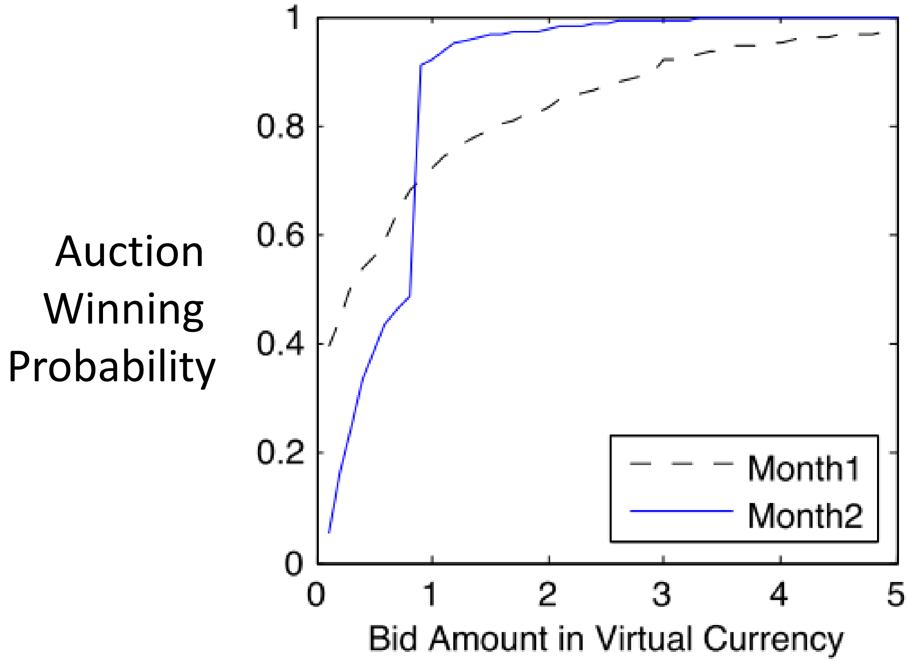


Bidding Strategy in Practice: A Quantitative Perspective



Bid Landscape Forecasting





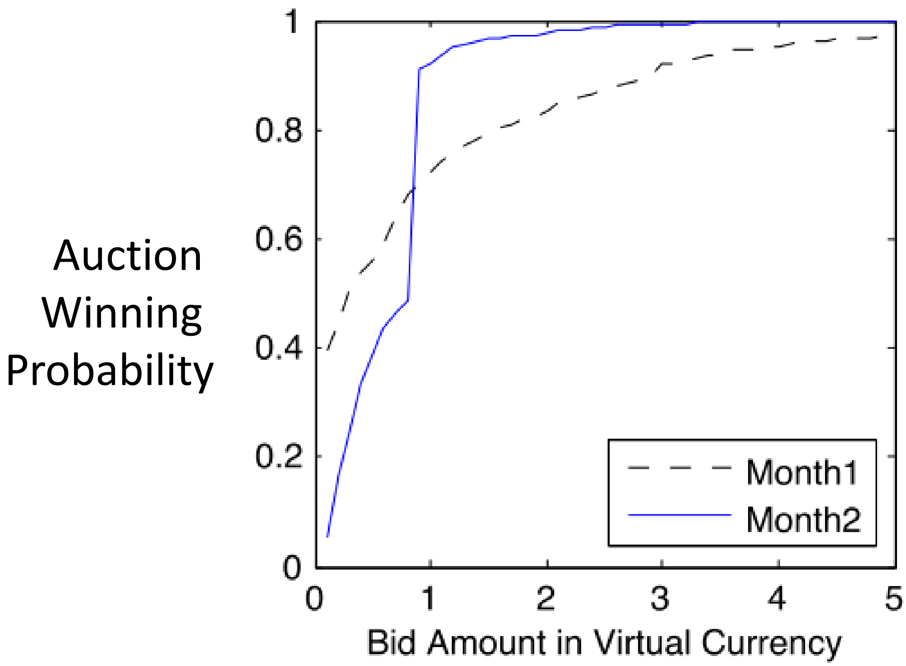
Win probability:

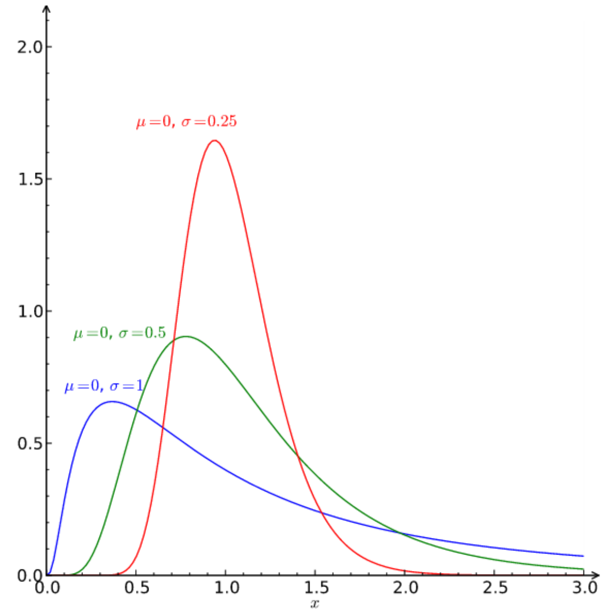
$$w(b) = \int_{z=0}^{b} p(z)dz$$

Expected cost:

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

Bid Landscape Forecasting





- Log-Normal Distribution

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

[Cui et al. Bid Landscape Forecasting in Online Ad Exchange Marketplace. KDD 11]

Bid Landscape Forecasting

- Price Prediction via Linear Regression

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

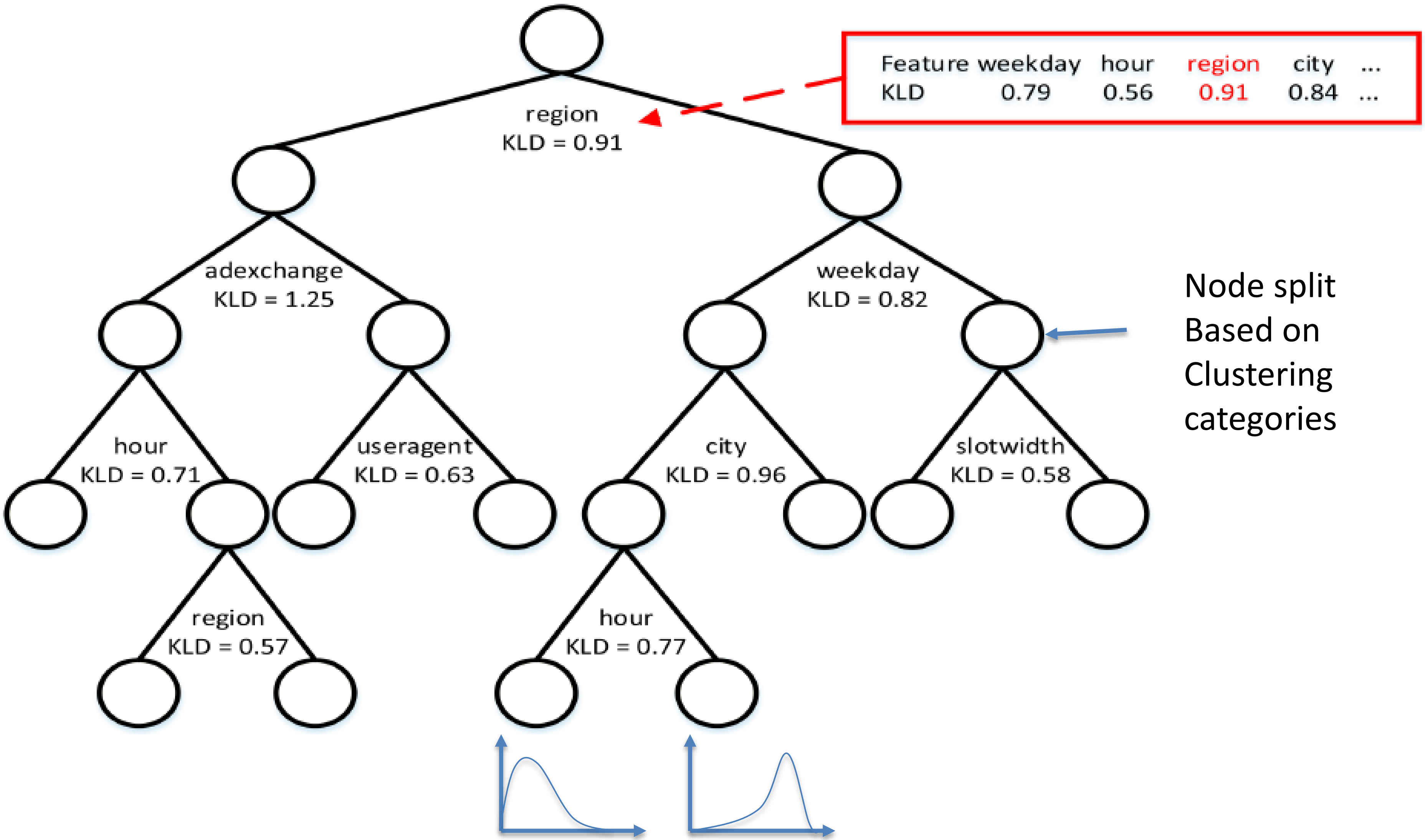
  – Modelling censored data in lost bid requests

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

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

[Wu et al. Predicting Winning Price in Real Time Bidding with Censored Data. KDD 15]

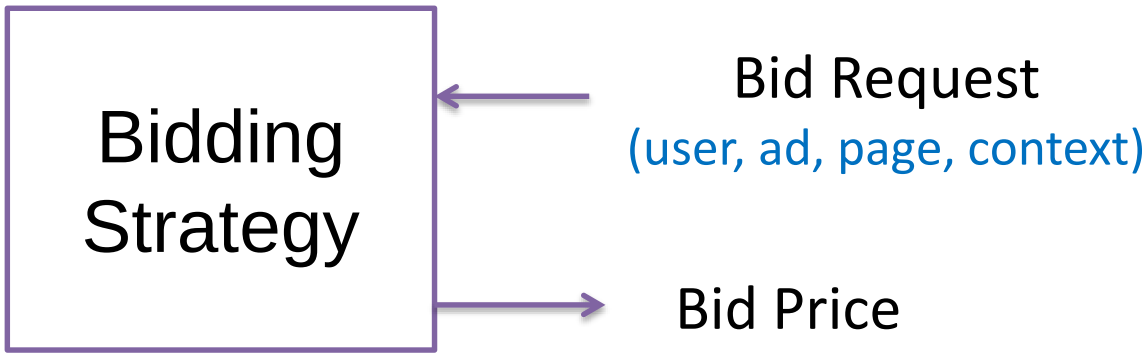
Survival Tree Models



[Yuchen Wang et al. Functional Bid Landscape Forecasting for Display Advertising. ECMLPKDD 2016]

Bidding Strategies

- How much to bid for each bid request?



- Bid to optimise the KPI with budget constraint

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

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

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

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

Truth-telling Bidding Strategies

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

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

[Chen et al. Real-time bidding algorithms for performance-based display ad allocation. KDD 11]

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

[Chen et al. Real-time bidding algorithms for performance-based display ad allocation. KDD 11]

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

$$\max_{\text{bidding strategy}} \quad \text{KPI}$$
$$\text{subject to} \quad \text{cost} \leq \text{budget}$$

[Perlich et al. Bid Optimizing and Inventory Scoring in Targeted Online Advertising. KDD 12]

ORTB Bidding Strategies

- Direct functional optimisation

winning function CTR

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

bidding function

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

Est. volume cost upperbound

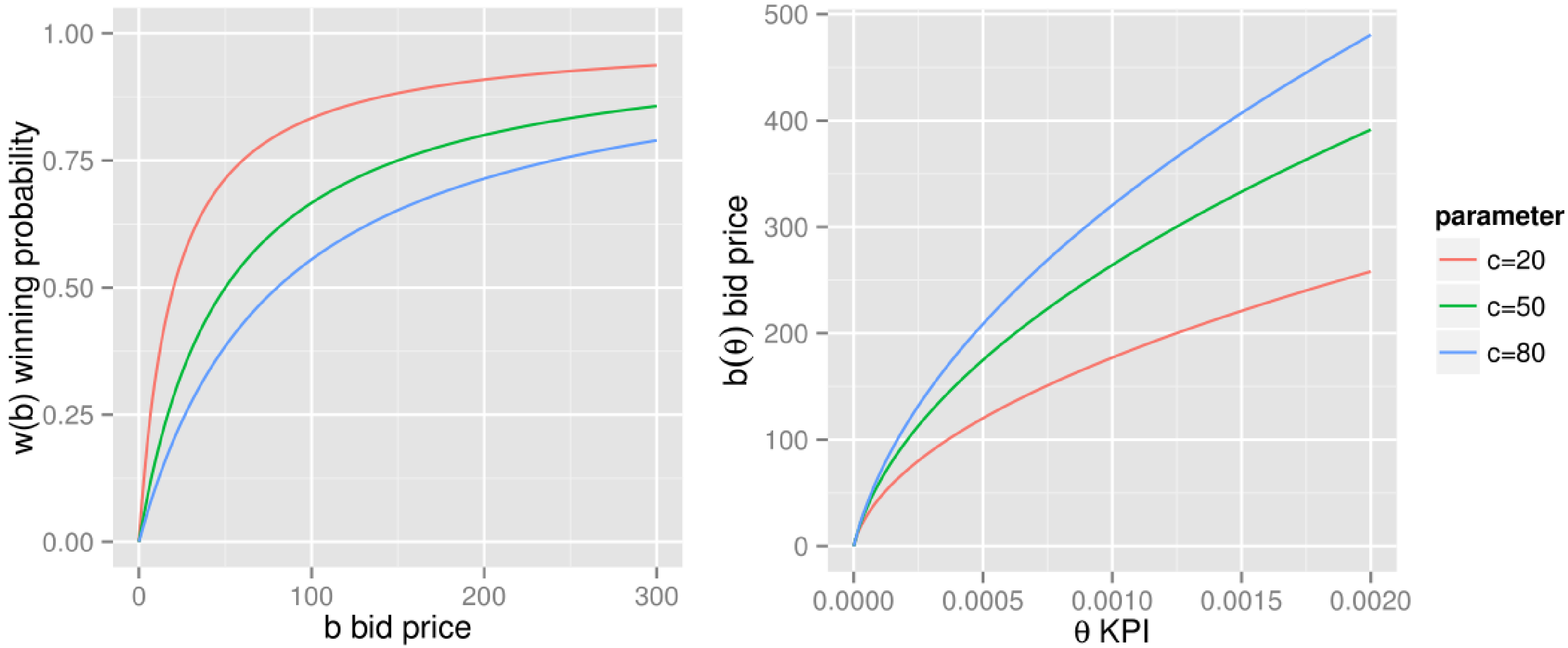
- Solution: Calculus of variations

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

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

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

Optimal Bidding Strategy Solution



(a) Winning function 1. (b) Bidding function 1.

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

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

Unbiased Optimisation

- Bid optimization on 'true' distribution

$$\arg\max_{b()} \ T \int_{\boldsymbol{x}} f(\boldsymbol{x}) w(b(f(\boldsymbol{x}))) p_x(\boldsymbol{x}) d\boldsymbol{x}$$

$$\text{subject to} \ \ T \int_{\boldsymbol{x}} b(f(\boldsymbol{x})) w(b(f(\boldsymbol{x}))) p_x(\boldsymbol{x}) d\boldsymbol{x} = B$$

- Unbiased bid optimization on biased distribution

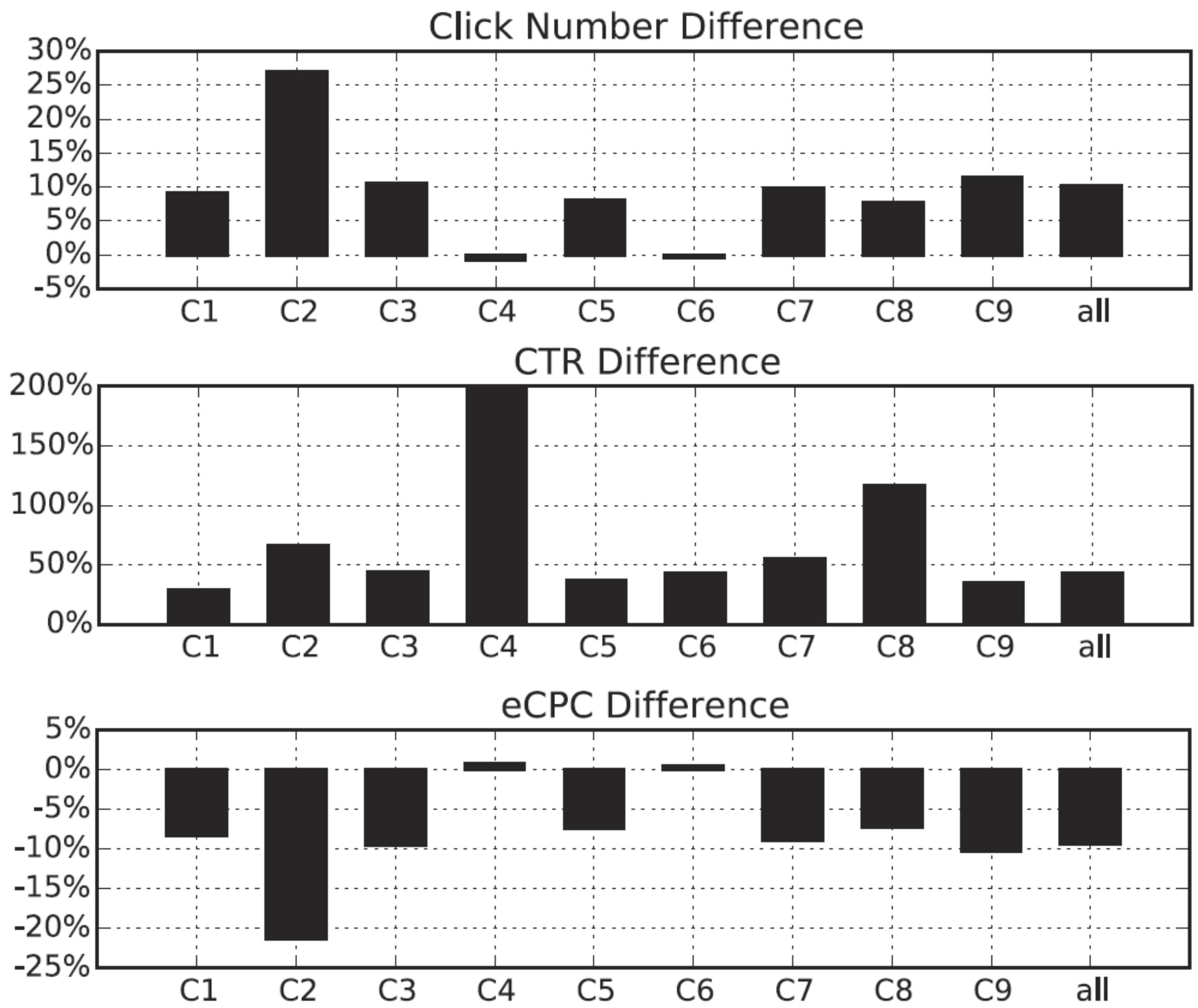
$$\arg\max_{b()} \ T \int_{\boldsymbol{x}} f(\boldsymbol{x}) w(b(f(\boldsymbol{x}))) \frac{q_x(\boldsymbol{x})}{w(b_{\boldsymbol{x}})} d\boldsymbol{x}$$

$$\text{subject to} \ \ T \int_{\boldsymbol{x}} b(f(\boldsymbol{x})) w(b(f(\boldsymbol{x}))) \frac{q_x(\boldsymbol{x})}{w(b_{\boldsymbol{x}})} d\boldsymbol{x} = B$$

[Zhang et al. Bid-aware Gradient Descent for Unbiased Learning with Censored Data in Display Advertising. KDD 2016.]

Unbiased Bid Optimisation

A/B Testing on Yahoo! DSP.



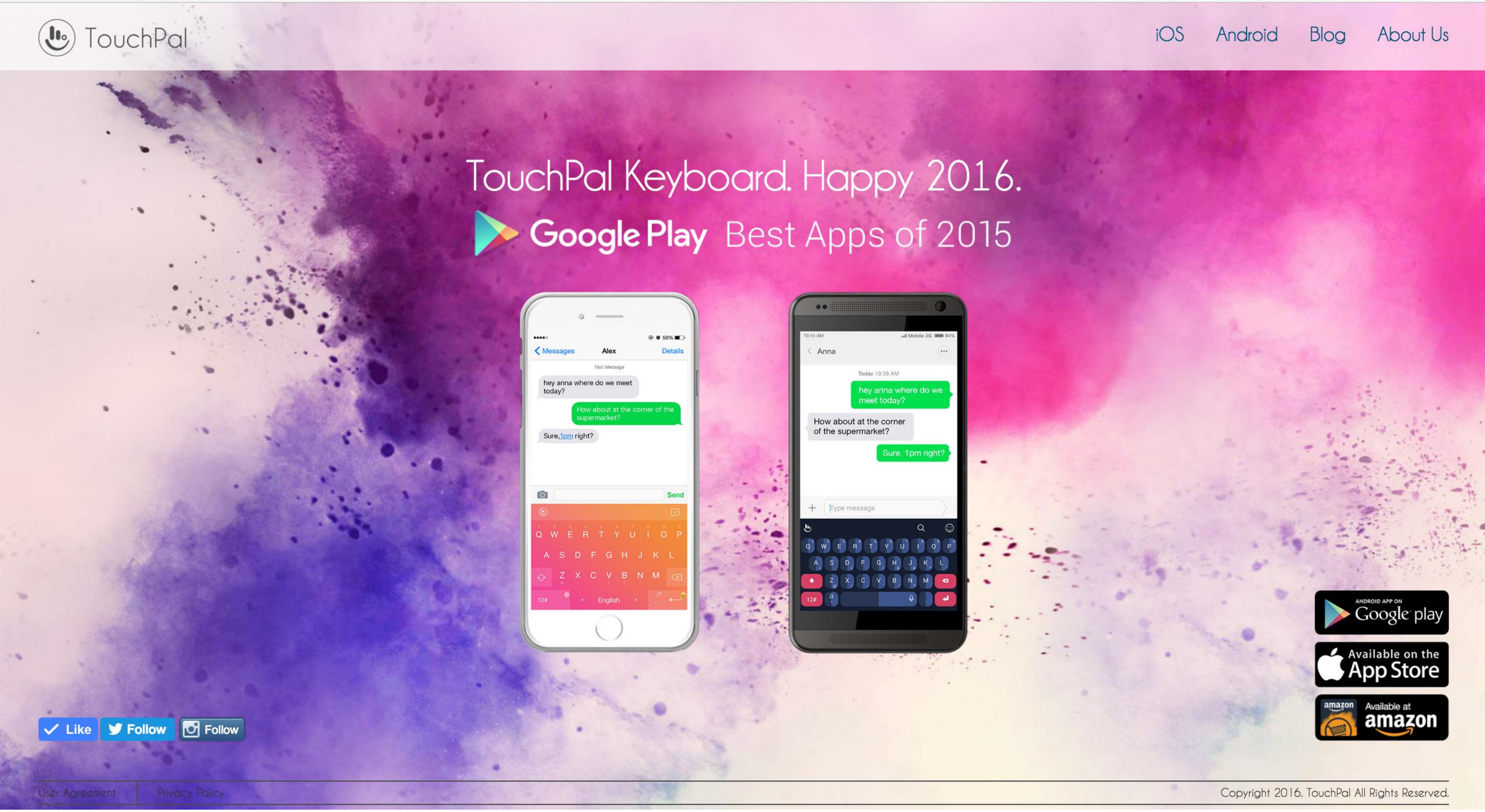
[Zhang et al. Bid-aware Gradient Descent for Unbiased Learning with Censored Data in Display Advertising. KDD 2016.]

That's the first half of the tutorial! Questions?

Part 2

Speaker: Jian Xu, TouchPal Inc.

(jian.xu AT cootek.cn)

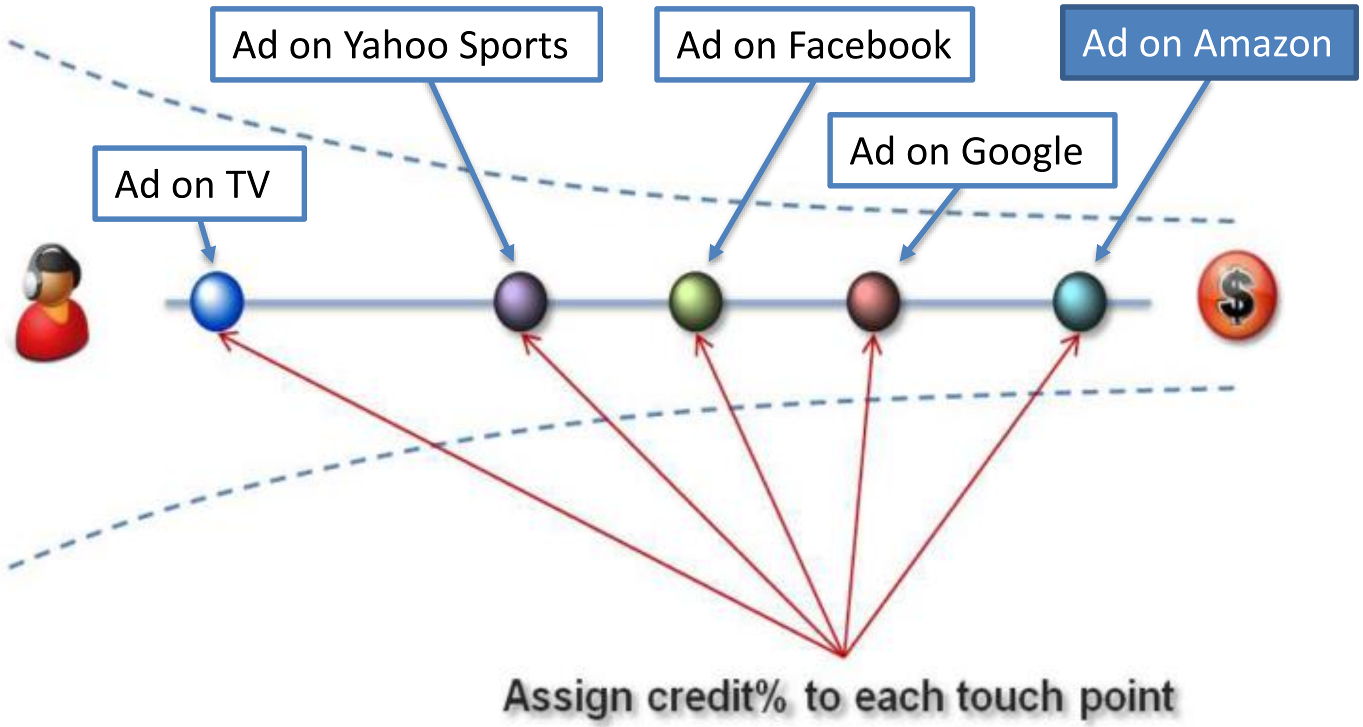


usjobs@cootek.cn

Table of contents

Table of contents

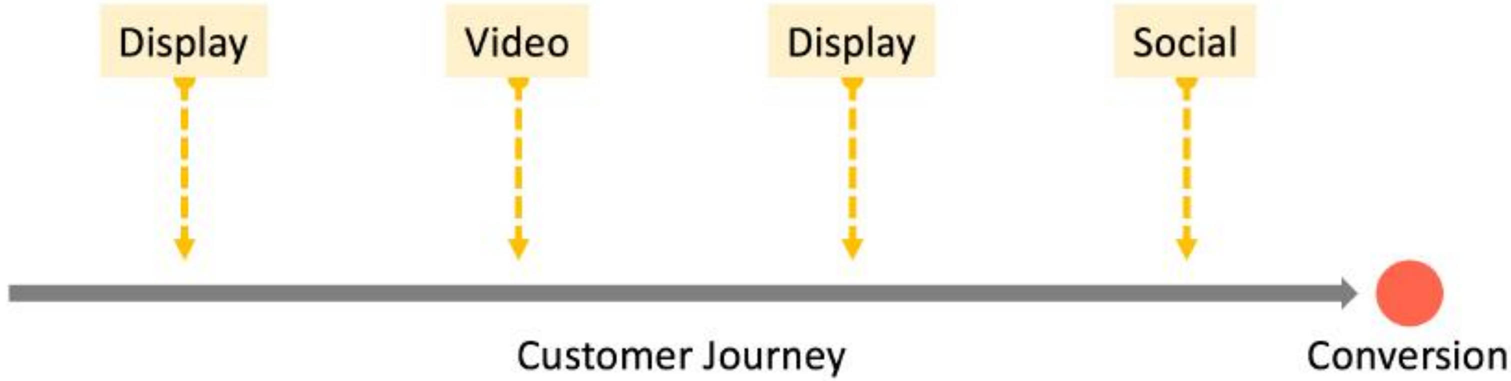
Conversion Attribution



- 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]

Rule-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
  - It should be built based on ad touch and conversion data of a campaign
- Interpretability
  - Generally accepted by all the parties

[Dalessandro et al. Casually Motivated Attribution for Online Advertising. ADKDD 11]

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 model and record the feature weights
- Average the weights of a feature

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

A Probabilistic Attribution Model

- Conditional probabilities

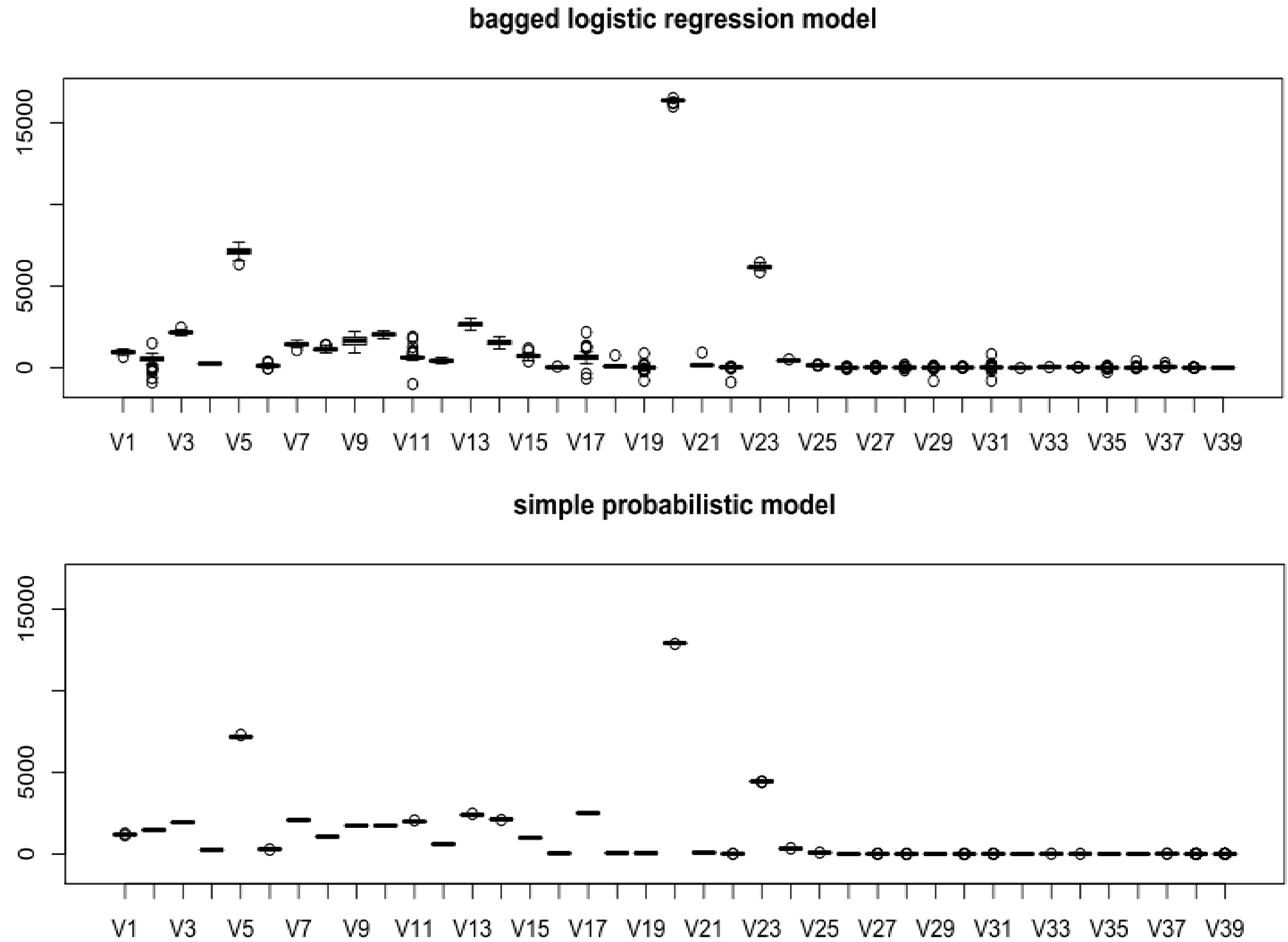
$$P(y|x_i) = \frac{N_{positive}(x_i)}{N_{positive}(x_i) + N_{negative}(x_i)}$$

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

- Attributed contribution (not-normalized)

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

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



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

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]

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

- A more generalized and data-driven model

  [Dalessandro et al. Causally Motivated Attribution for Online Advertising. ADKDD 11]

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

  – $w_{S,i}$ is the probability that the ad touch sequence begins with S, x_i

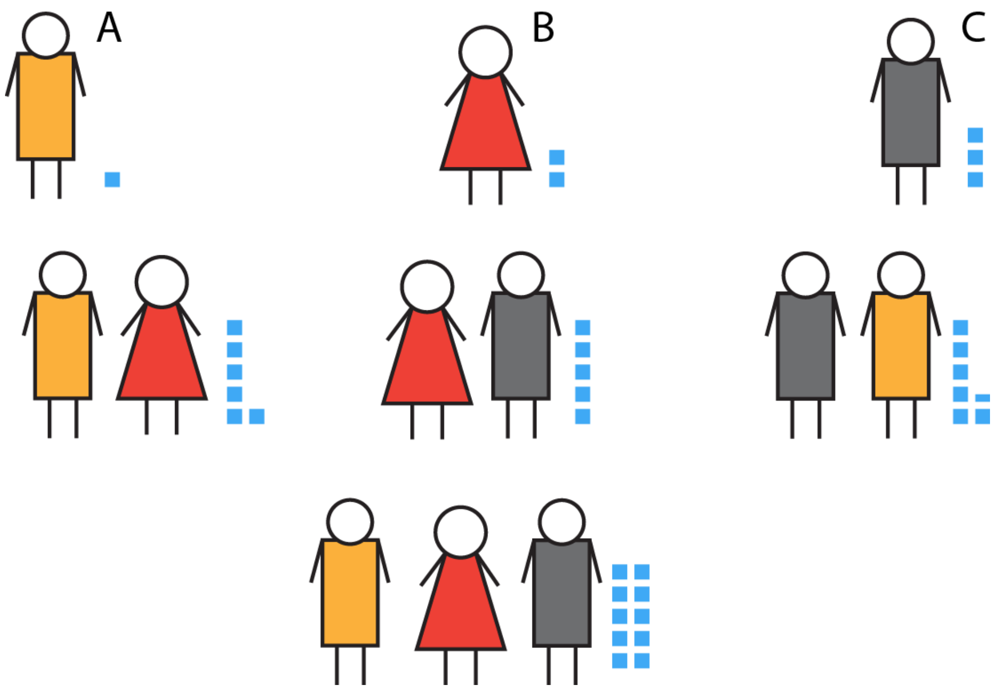
Attribution Comparison: LTA vs MTA

		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%

[Dalessandro et al. Casually Motivated Attribution for Online Advertising. ADKDD 11]

Shapley Value based Attribution

- Coalitional 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

- Coalitional game
  - v is the conversion rate of different subset of publishers
  - The Shapley value of publisher i is

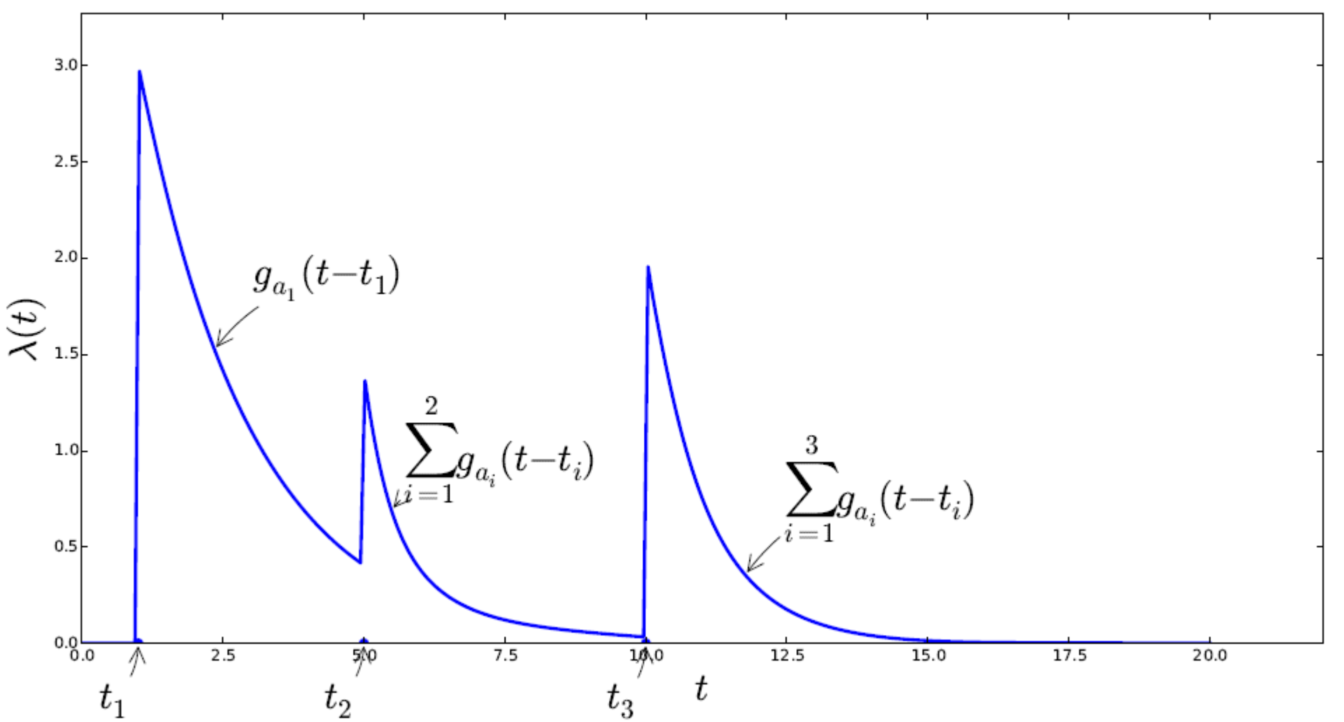
$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!} \left(\boxed{v(S \cup \{i\})} - v(S)\right)$$

CVR of those touched by all the publishers in $S \cup \{i\}$

[Berman, Ron. Beyond the last touch: Attribution in online advertising." *Available at SSRN 2384211* (2013)]

Survival theory-based model

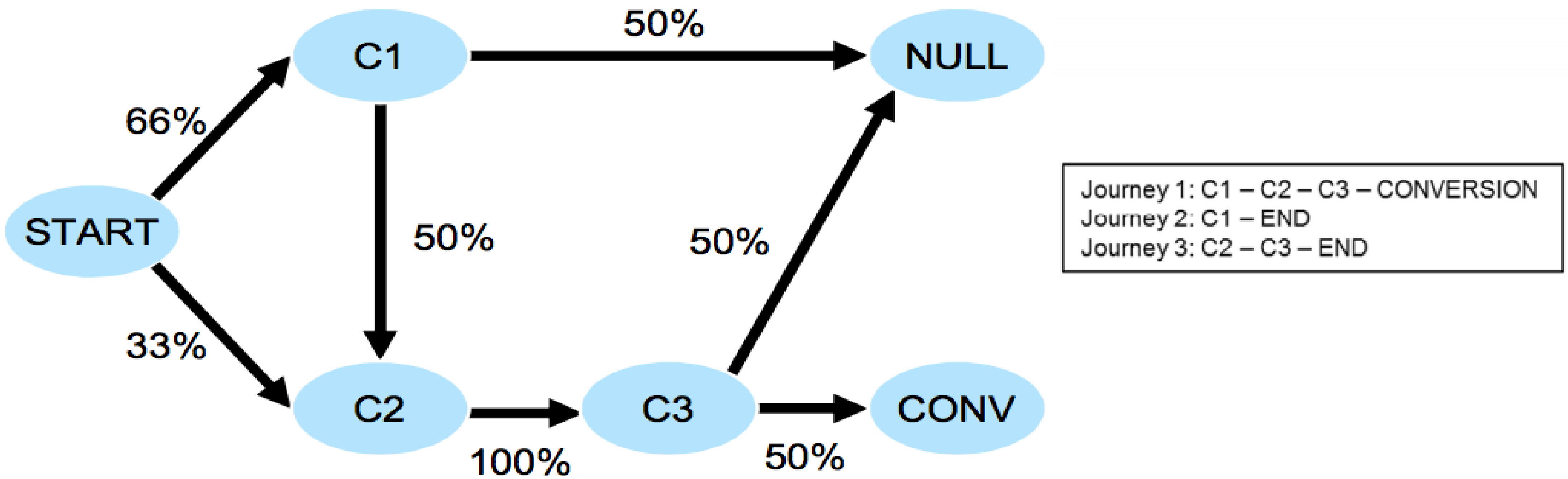
- Use *addictive* hazard functions to explicitly model:
  - the strength of influence, and
  - the time-decay of the influence



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

Markov graph-based approach

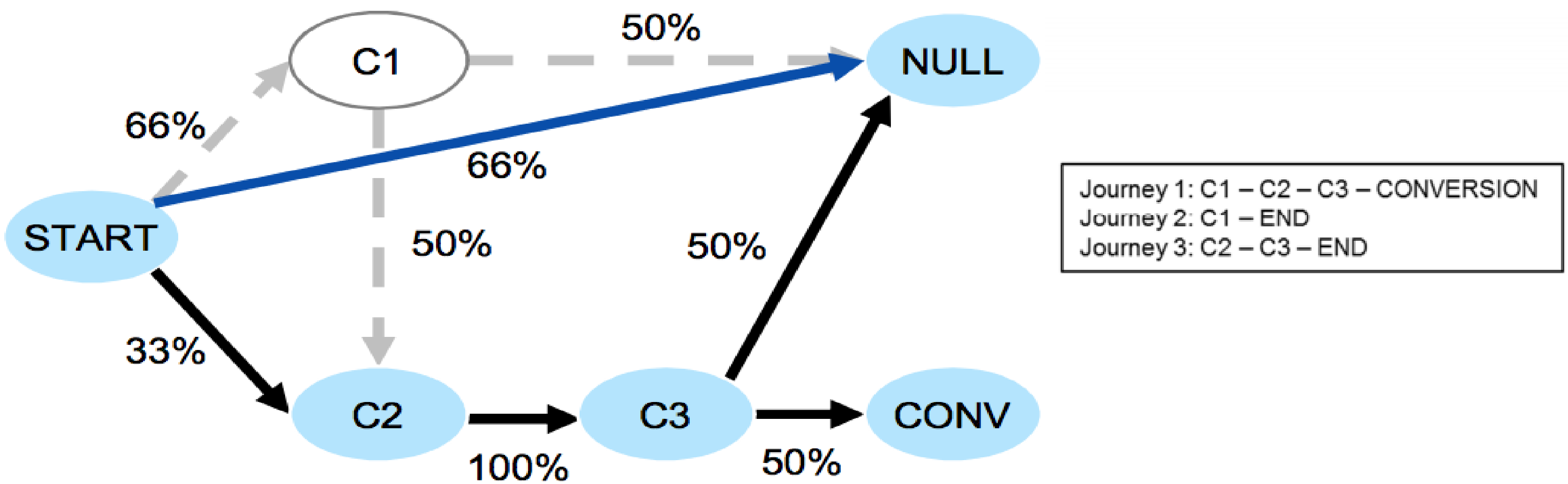
- Establish a graph from observed user journeys



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

Markov graph-based approach

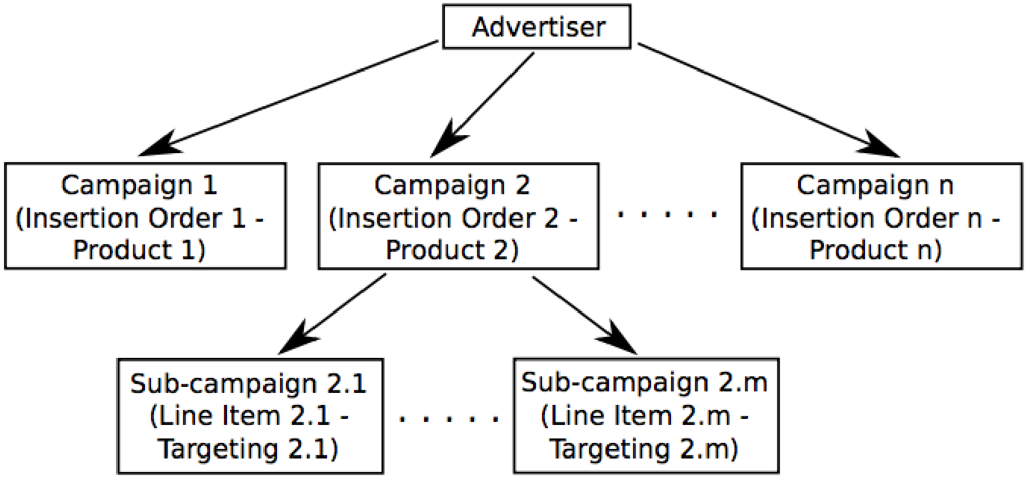
- Attribute based on probability change of reaching conversion state



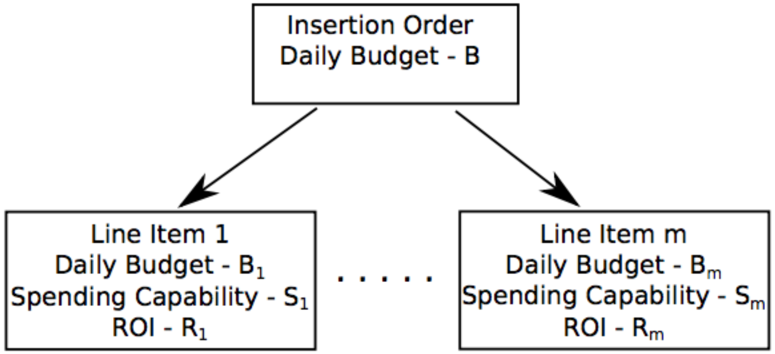
[Anderl et al. Mapping the customer journey: A graph-based framework for online attribution modeling. SSRN 2014]

MTA-based budget allocation

- Typical advertiser hierarchy



- Typical budget allocation scheme



[Geyik et al. Multi-Touch Attribution Based Budget Allocation in Online Advertising. ADKDD 14]

MTA-based budget allocation

- Estimate sub-campaign spending capability
  - New sub-campaign: assign a learning budget
  - Existing sub-campaign: assign an x% more budget
- Calculate ROI of each sub-campaign

$$\mathrm{ROI}_{l_i} = \frac{\sum_{\forall a_j} \boxed{p(l_i|a_j)}\, v(a_j)}{\text{Money spent by } l_i}$$

1 if l_i is the last touch point else 0 (LTA)

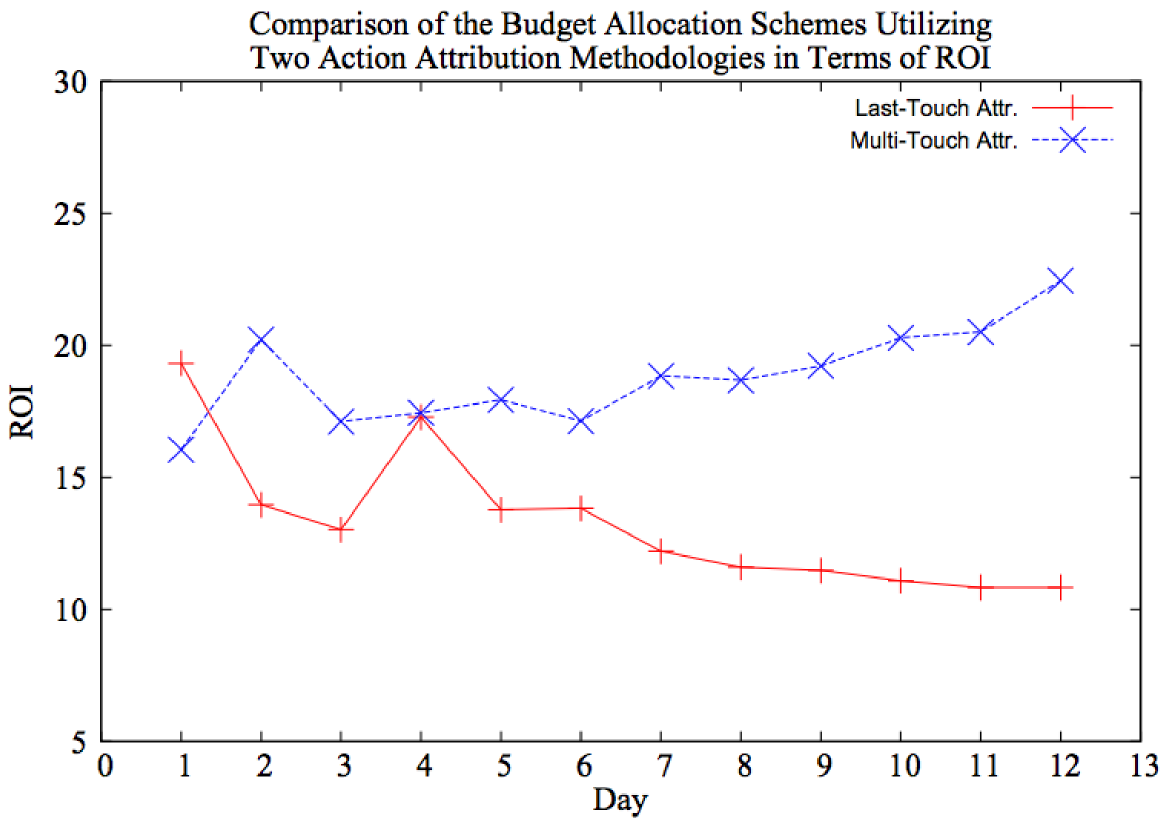
$$\frac{V(l_i)}{\Sigma_{l_k \in S_{a_j}} V(l_k)} \quad \text{(MTA)}$$

- Allocate budget in a cascade fashion

[Geyik et al. Multi-Touch Attribution Based Budget Allocation in Online Advertising. ADKDD 14]

MTA-based budget allocation

- Results on a real ad campaign



[Geyik et al. Multi-Touch Attribution Based Budget Allocation in Online Advertising. ADKDD 14]

Attribution and Bidding

- For CPA campaigns, *conventional* bidding strategy is to bid prop. to estimated action rate (a.k.a. conversion rate). Is that always correct?

A tiny example

Two users: a and b
AR_a: 0.04 if exposed to the ad, 0.03 if not;
AR_b: 0.02 if exposed to the ad, 0.001 if not.

If only one of them can be exposed to the ad, who will you select?

[Xu et al. Lift-Based Bidding in Ad Selection. AAAI 2016.]

Attribution and Bidding

A not-so-tiny example

Two users: a and b, campaign CPA: \$100
AR_a: 0.04 if exposed to the ad, 0.03 if not (lift: 0.01);
AR_b: 0.02 if exposed to the ad, 0.001 if not (lift: 0.019).
Bidder_1 bids prop. to AR assuming exposed: \$4 for a, \$2 for b;
Bidder_2 bids prop. to AR lift: \$2 for a, \$3.8 for b.
Incremental value from Bidder_1: 0.01 conversions;
Incremental value from Bidder_2: 0.19 conversions.
Expected attribution to Bidder_1: 0.04 conversions;
Expected attribution to Bidder_2: 0.02 conversions.

- *Prevalent bidding strategy does not optimize campaign performance;*
- *Bidders are not rewarded fairly.*

Rational DSPs for CPA advertisers

- DSP's perspective:
  - *Cost*: second price in the auction
  - *Reward*: CPA if (1) there is action, and (2) the action is attributed to it
  - A *rational DSP* will always bid

$$\text{bid} = \text{AR} \times \text{CPA} \times p(attribution|action)$$

In LTA, *p*(attribution|action) is always 1 for the last toucher. ***Therefore DSPs are bidding to maximize their chance to be attributed instead of maximizing conversions.***

Bidding in Multi-Touch Attribution

- Current bidding strategy (driven by LTA)

$$\text{bid} = \text{AR} \times \text{CPA}$$

- A new bidding strategy (driven by MTA)
  - If attribution is based on the AR lift

$$\Delta p = p(action|s_+(a)) - p(action|s)$$

$$\text{bid} = \boxed{\Delta p \times base\_bid}$$

Lift- based bidding

[Xu et al. Lift-Based Bidding in Ad Selection. AAAI 2016.]

Lift-based bidding

$$\text{bid} = \Delta p \times base\_bid$$

- Estimating action rate lift
  - Learn a *generic* action prediction model $\hat{P}$ on top of features extracted from *user-states* $F(s)$
  - Then action rate lift can be estimated by

$$\widehat{\Delta p} = \hat{P}(action|F(s_+(a))) - \hat{P}(action|F(s))$$

- Deriving the base_bid $\beta = \frac{\bar{p}}{\Delta p} \times \text{CPA}$

[Xu et al. Lift-Based Bidding in Ad Selection. AAAI 2016.]

Lift-based bidding

Value-based bidding vs. lift-based bidding - Advertiser's perspective

Adv	Value-based bidding			Lift-based bidding			Action lift	Lift-over-lift
	# imps	# actions	Action lift (vs "No bid")	# imps	# actions	Action lift (vs "no bid")		
1	53,396	714	11.2%	59,703	826	28.7%	13.6%	**156%**
2	298,333	896	8.9%	431,637	980	19.1%	9.4%	**115%**
3	11,048,583	1,477	2.7%	11,483,360	1509	4.9%	2.2%	**82%**
4	3,915,792	2,016	6.6%	4,368,441	2,471	30.6%	22.6%	**367%**
5	6,015,322	6,708	19.6%	8,770,935	8,291	47.8%	23.6%	**144%**

Value-based bidding vs. lift-based bidding - DSP's perspective.

Adv	Value-based bidding			Lift-based bidding			**Inventory-cost diff**	**Cost-per-imp diff**
	# imps	# attrs	Inventory cost	# imps	# attrs	Inventory cost		
1	53,396	50	$278.73	59,703	50	$300.31	**7.7%**	**-3.6%**
2	298,333	80	$1,065.05	431,637	80	$1,467.57	**37.8%**	**-4.8%**
3	11,048,583	240	$25,522.22	11,483,360	240	$25,837.56	**1.2%**	**-2.6%**
4	3,915,792	200	$10,846.74	4,368,441	200	$11,183.21	**3.1%**	**-7.6%**
5	6,015,322	500	$19,296.51	8,770,935	500	$23,501.90	**21.8%**	**-16.5%**

[Xu et al. Lift-Based Bidding in Ad Selection. AAAI 2016.]

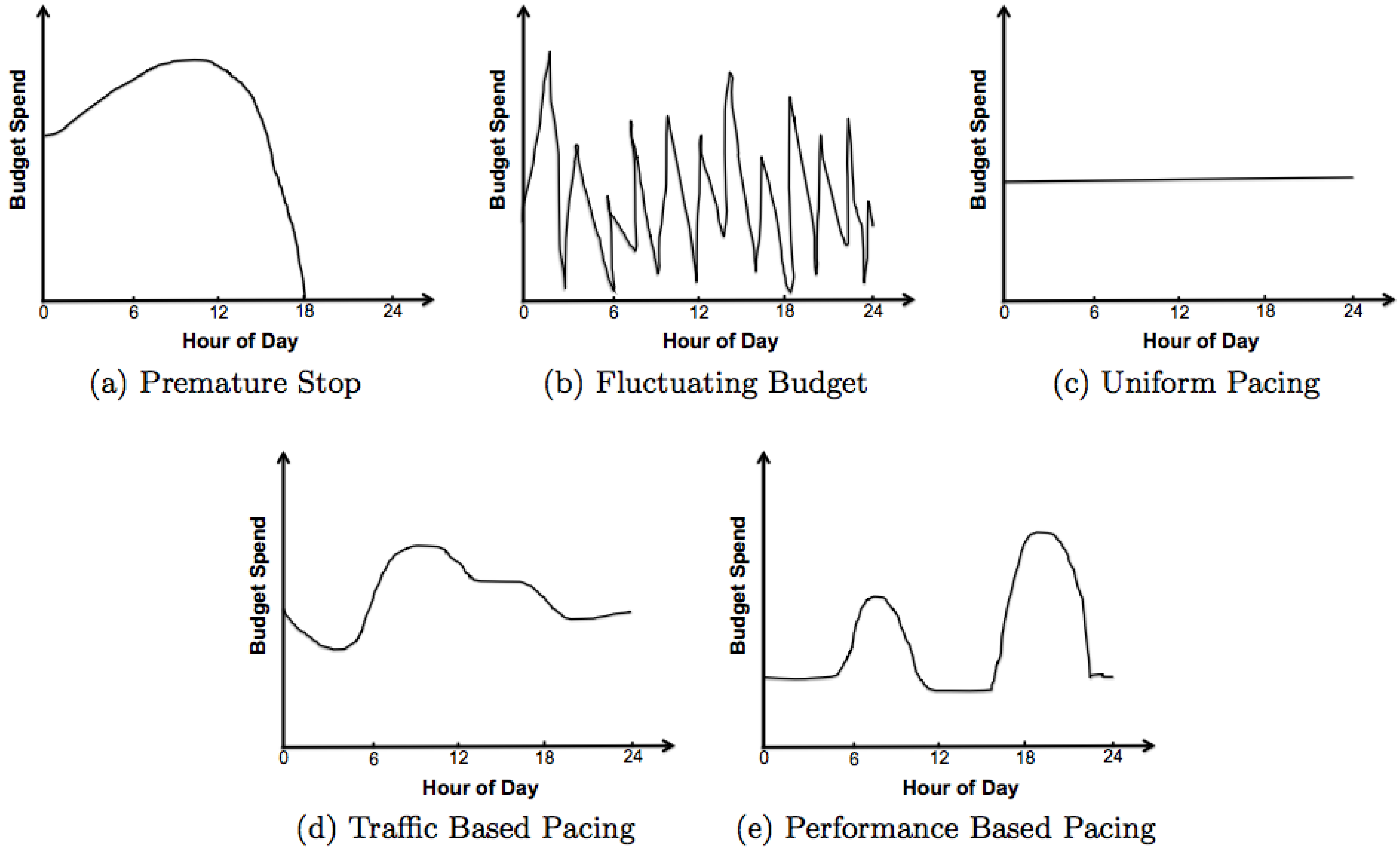
Table of contents

Pacing Control

- *Budget pacing control* helps advertisers to define and execute how their budget is spent over the time.

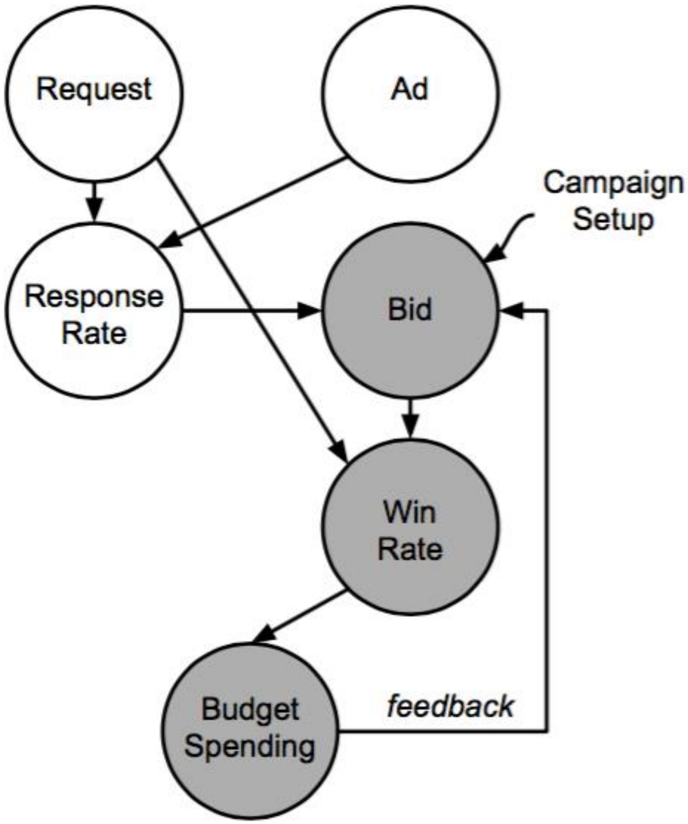
- Why?
  - Avoid premature campaign stop, overspending and spending fluctuations.
  - Reach a wider range of audience
  - Build synergy with other marketing campaigns
  - Optimize campaign performance

Examples

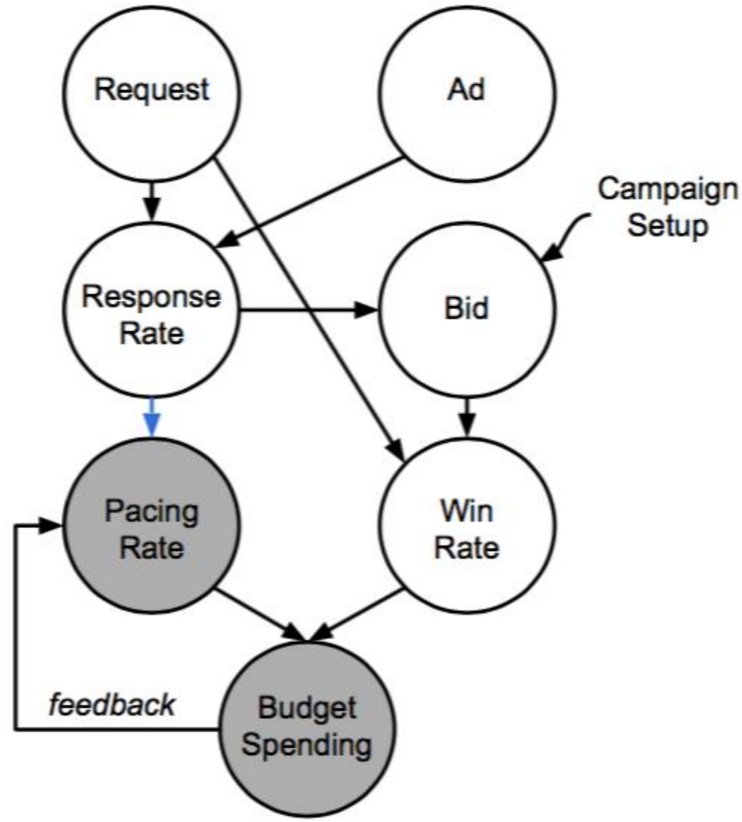


[Lee et al. Real Time Bid Optimization with Smooth Budget Delivery in Online Advertising. ADKDD 13]

Two streams of approaches



Bid modification

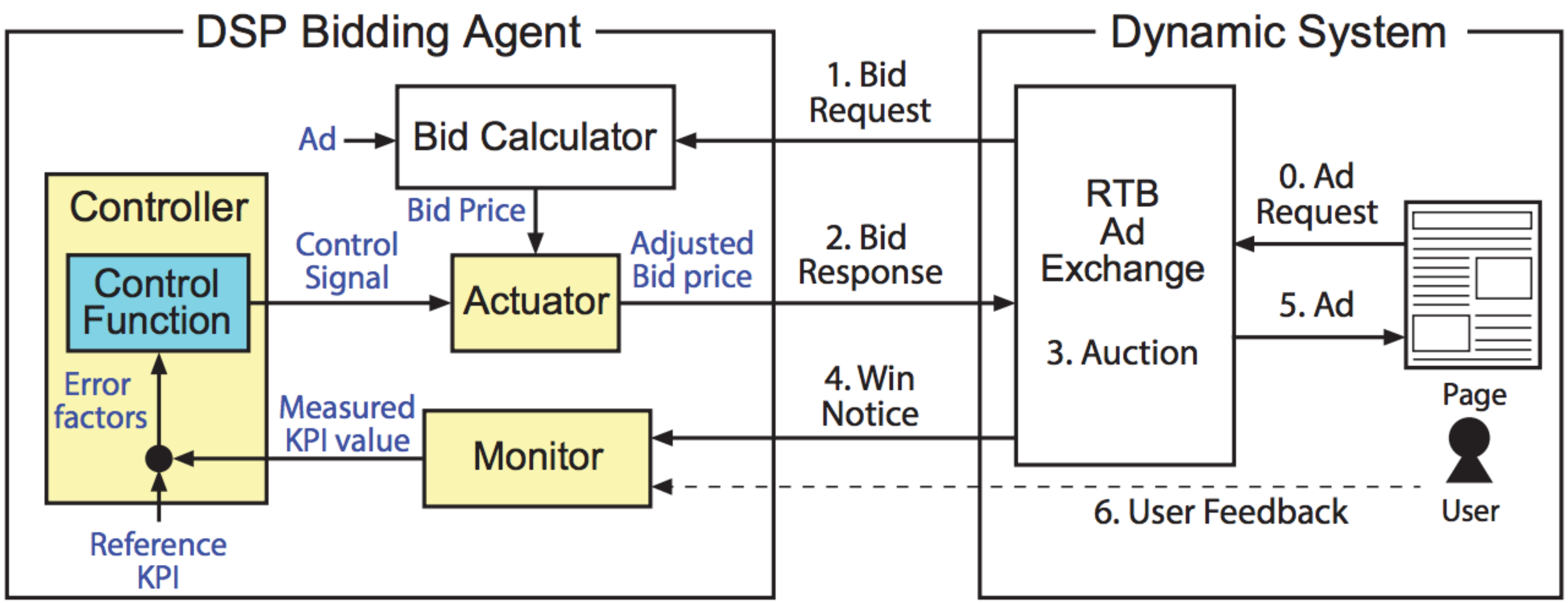


Probabilistic throttling

[Xu et al. Smart Pacing for Effective Online Ad Campaign Optimization. KDD 2015.]

Bid modification with PID controller

- Add a *monitor*, a *controller* and an *actuator* module into the bidding system
- Achieve reference KPI (e.g. eCPC) by bid modification



[Zhang et al. Feedback Control of Real-Time Display Advertising. WSDM 2016.]

Bid modification with PID controller

- Current control signal is calculated by PID controller

$$e(t_k) = \boxed{x_r} - \boxed{x(t_k)},$$

Reference KPI

Actual KPI value

$$\phi(t_{k+1}) \leftarrow \lambda_P e(t_k) + \lambda_I \sum_{j=1} e(t_j)\Delta t_j + \lambda_D \frac{\Delta e(t_k)}{\Delta t_k}$$

- Bid price is adjusted by taking into account current control signal

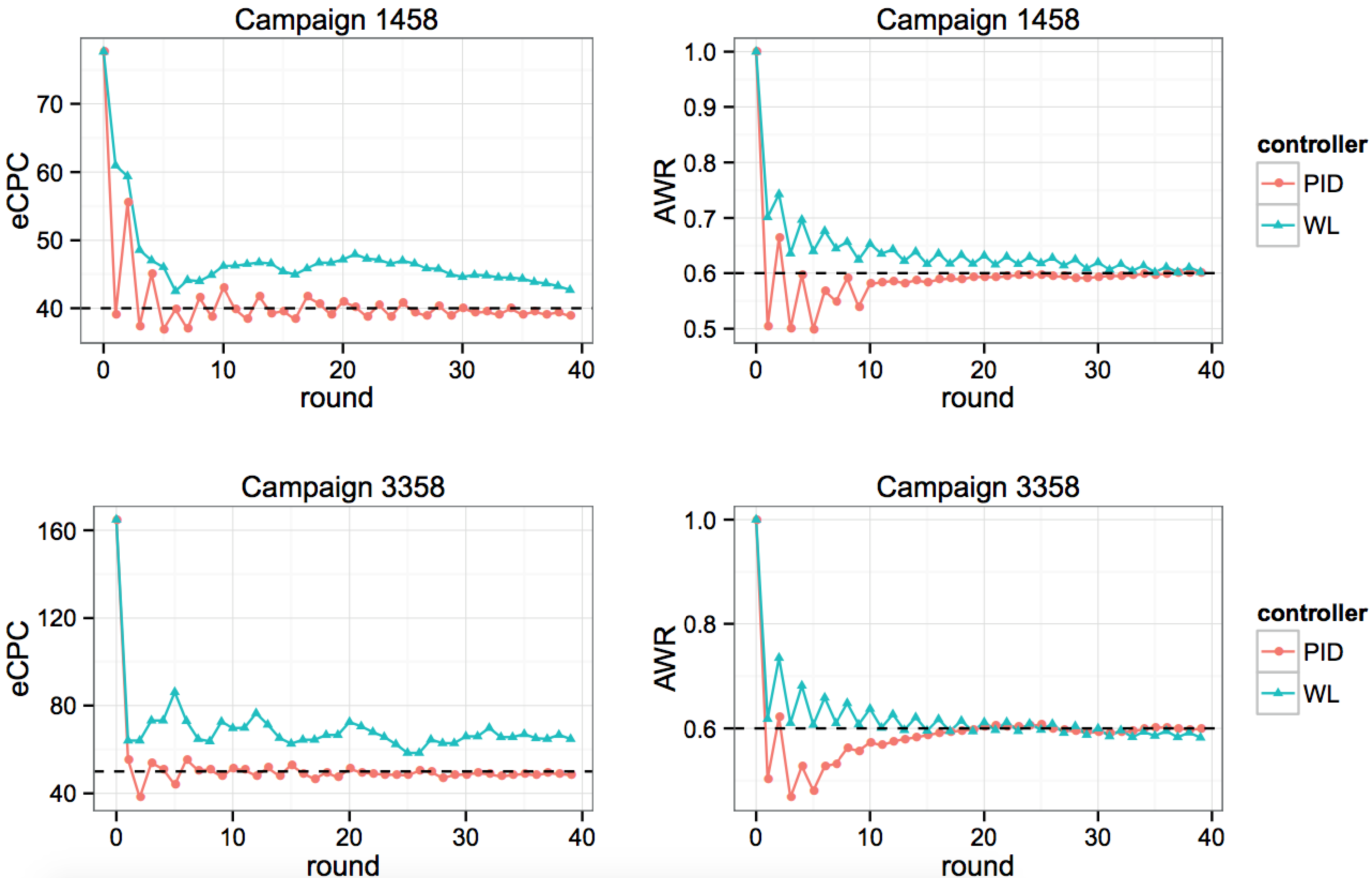
$$b_a(t) = b(t)\exp\{\boxed{\phi(t)}\}$$

The control signal

- A baseline controller: Water-level controller

$$\phi(t_{k+1}) \leftarrow \phi(t_k) + \gamma(x_r - x(t_k))$$

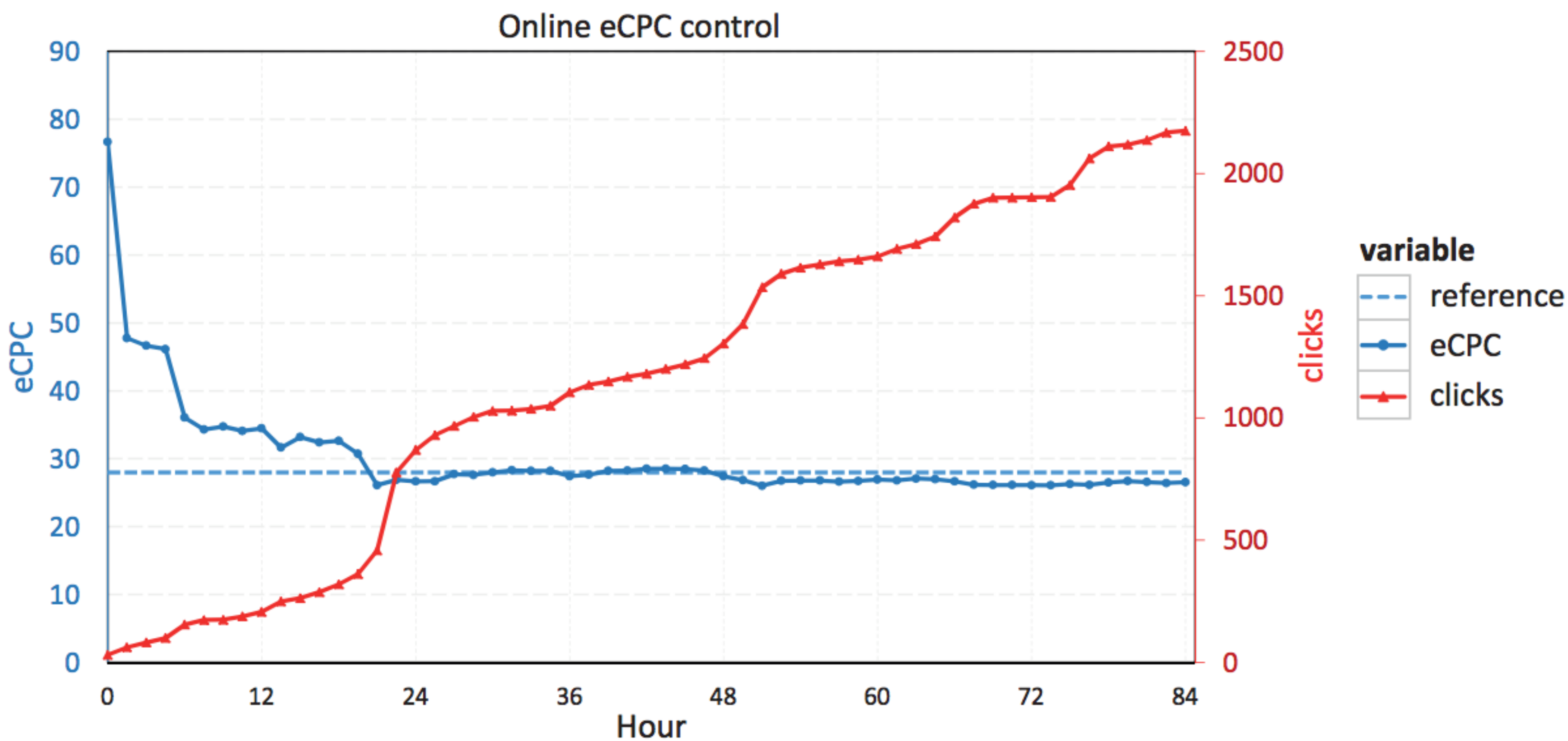
[Zhang et al. Feedback Control of Real-Time Display Advertising. WSDM 2016.]



Campaign 1458
eCPC
round
Campaign 1458
AWR
round
controller
PID
WL
Campaign 3358
eCPC
round
Campaign 3358
AWR
round
controller
PID
WL

Bid modification with PID controller

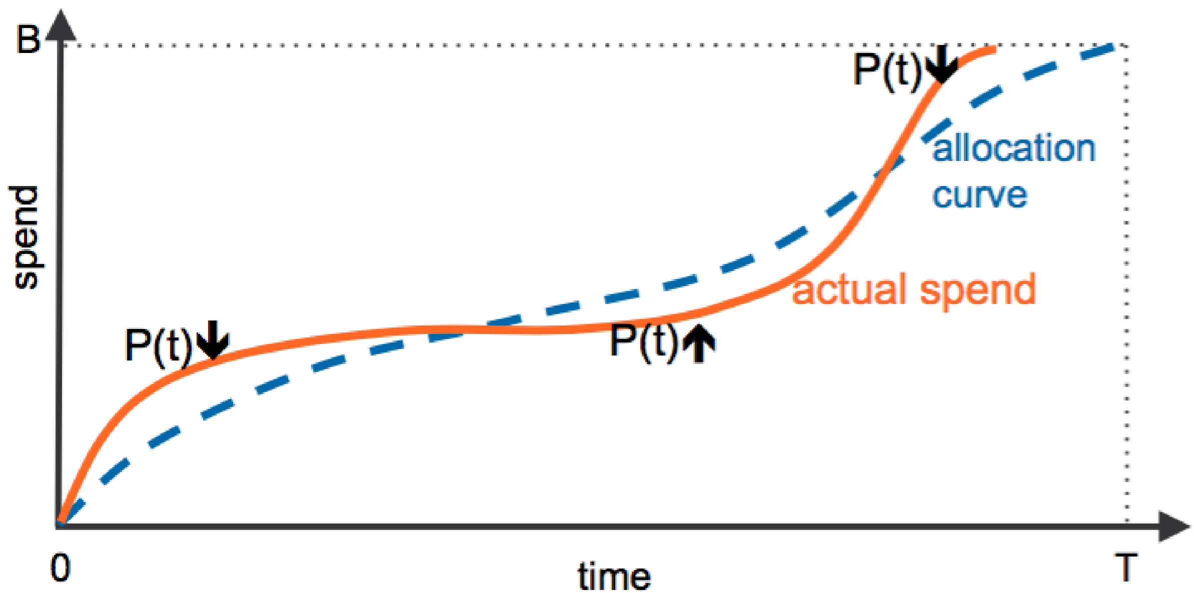
- Online eCPC control performance of a mobile game campaign



[Zhang et al. Feedback Control of Real-Time Display Advertising. WSDM 2016.]

Probabilistic throttling with conventional feedback controller

- P(t): pacing-rate at time slot t
- Leverage a *conventional feedback controller*:
  - P(t)=P(t–1)*(1–R) if budget spent > allocation
  - P(t)=P(t–1)*(1+R) if budget spent < allocation



[Agarwal et al. Budget Pacing for Targeted Online Advertisements at LinkedIn. KDD 2014.]

Probabilistic throttling with adaptive controller

- Leverage an *adaptive controller*

Desired spending in the next time-slot

$$\begin{aligned}
&\mathbf{pacing_rate}(t+1) \\
&= \mathbf{pacing_rate}(t)\frac{\boxed{\mathbf{s}(t+1)}}{\mathbf{s}(t)}\frac{\mathbf{reqs}(t)}{\boxed{\mathbf{reqs}(t+1)}}\frac{\mathbf{win_rate}(t)}{\boxed{\mathbf{win_rate}(t+1)}} \\
&= \mathbf{pacing_rate}(t)\frac{b_{t+1}}{\mathbf{s}(t)}\frac{\mathbf{reqs}(t)}{\mathbf{reqs}(t+1)}\frac{\mathbf{win_rate}(t)}{\mathbf{win_rate}(t+1)}
\end{aligned}$$

Forecasted request volume and bid win rate in the next time-slot

b_{t+1} is the desired spend (allocated) at time slot t+1. Different desired spending patterns can incur different calculation.

[Lee et al. Real Time Bid Optimization with Smooth Budget Delivery in Online Advertising. ADKDD 13]

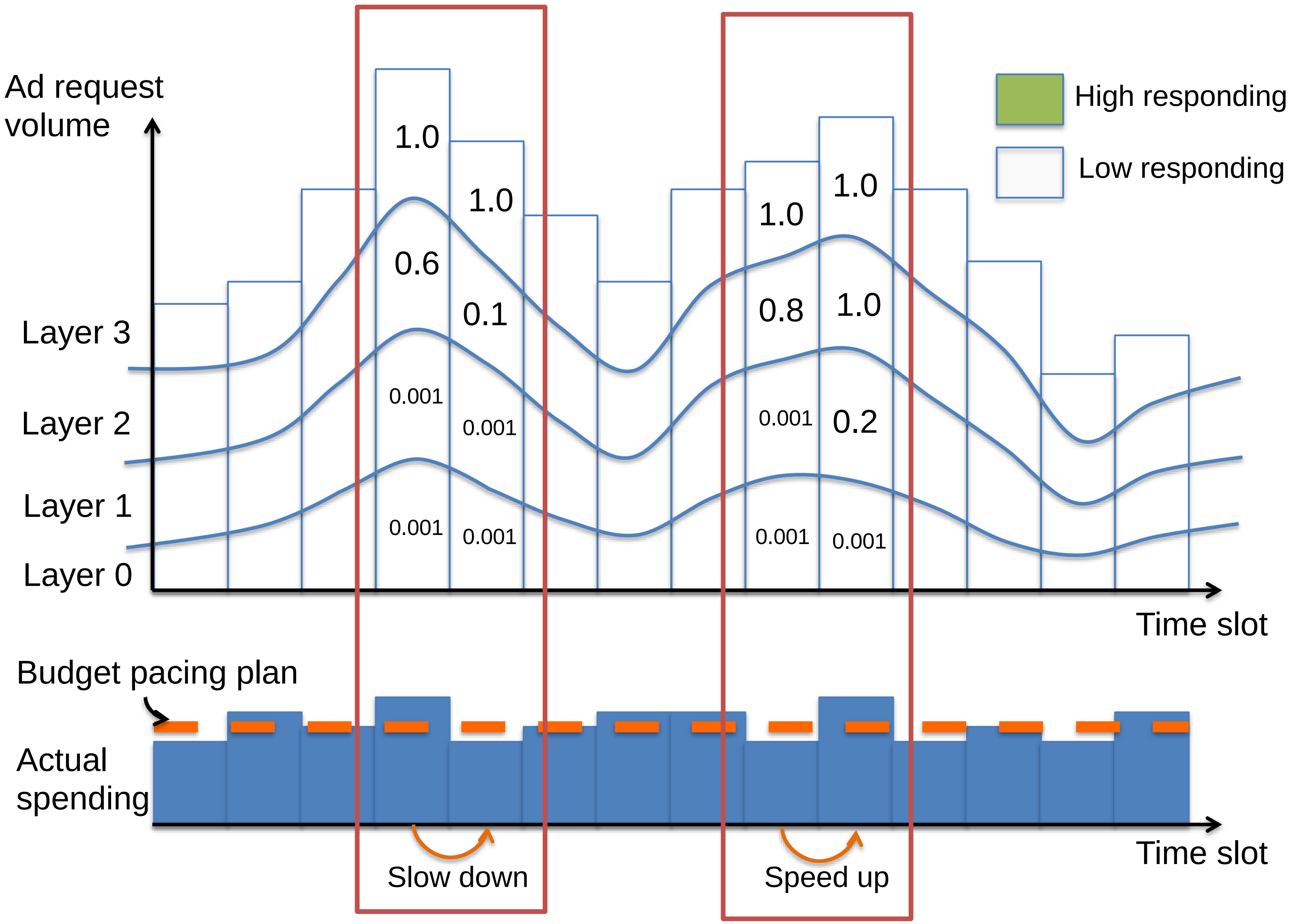
Pacing control for campaign optimization

- Campaign optimization objectives:
  - Reach delivery and performance goals
    - *Branding campaigns*: Spend out budget > Campaign performance (e.g., in terms of eCPC or eCPA)
    - *Performance campaigns*: Meet performance goal > Spend as much budget as possible.
  - Execute the budget pacing plan
  - Reduce creative serving cost

Can we achieve all these objectives by pacing control?

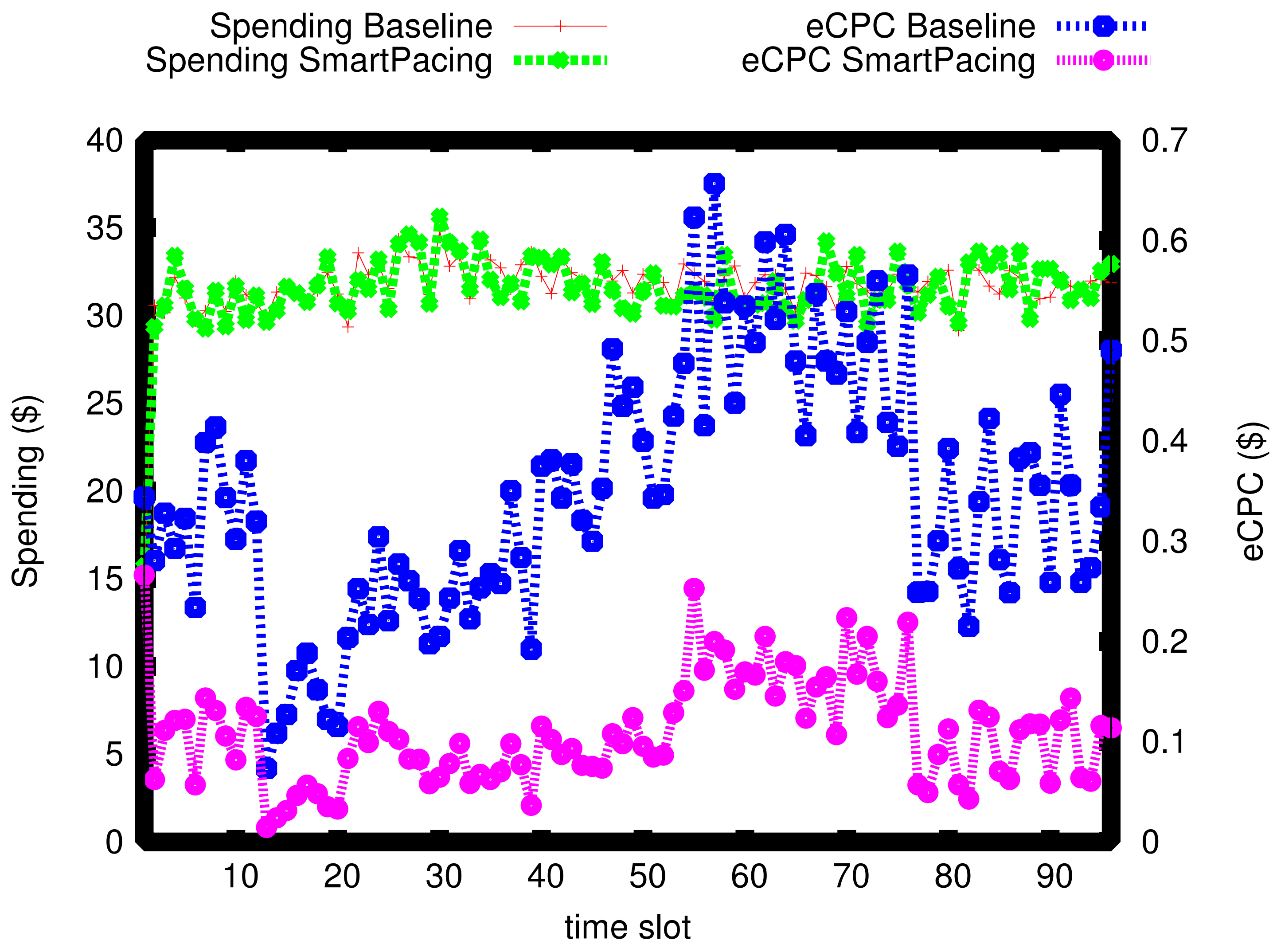
[Xu et al. Smart Pacing for Effective Online Ad Campaign Optimization. KDD 2015.]

Smart pacing

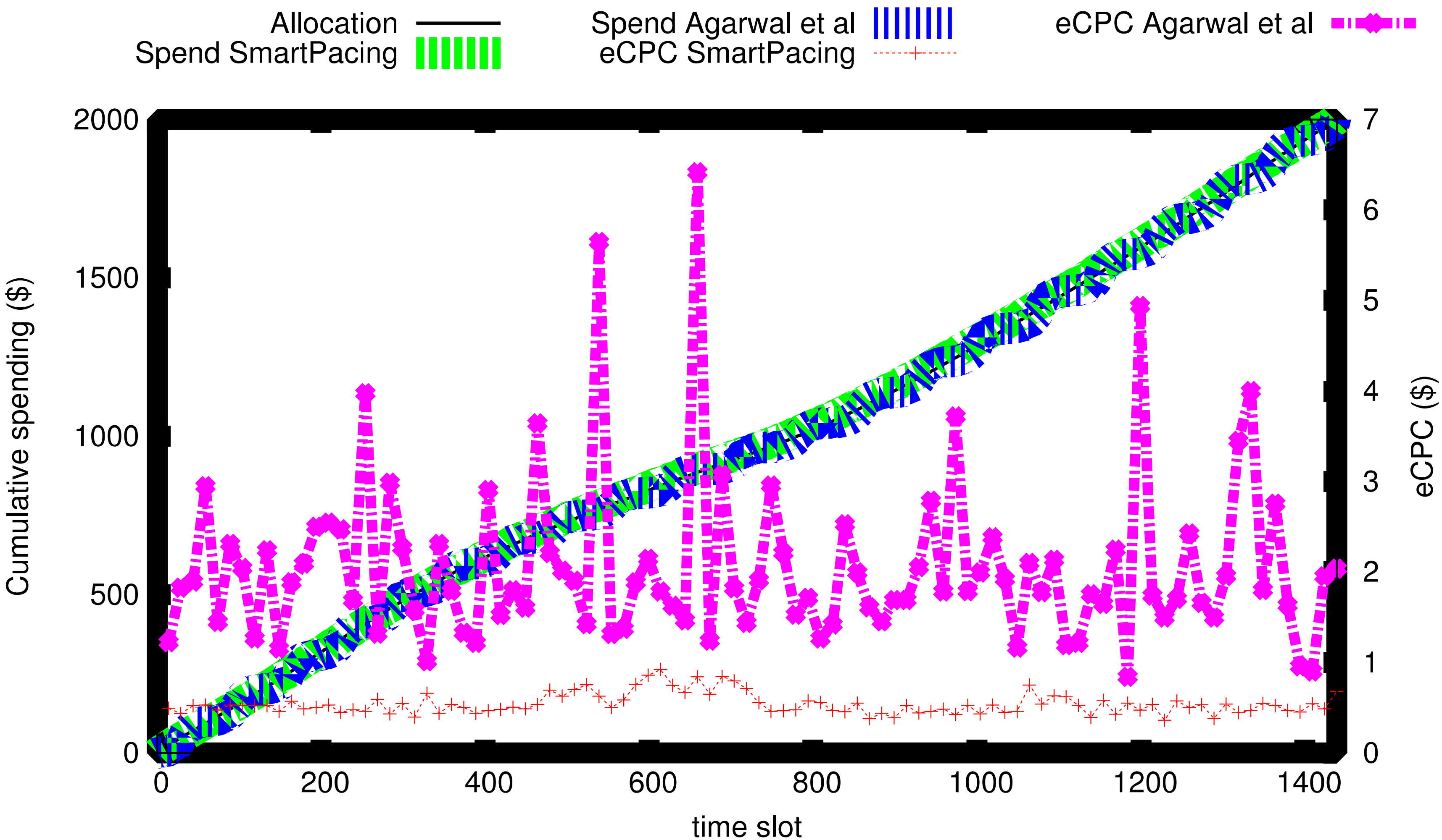


[Xu et al. Smart Pacing for Effective Online Ad Campaign Optimization. KDD 2015.]

Smart pacing performance



Smart pacing vs conventional feedback controller



Smart pacing vs conventional feedback controller

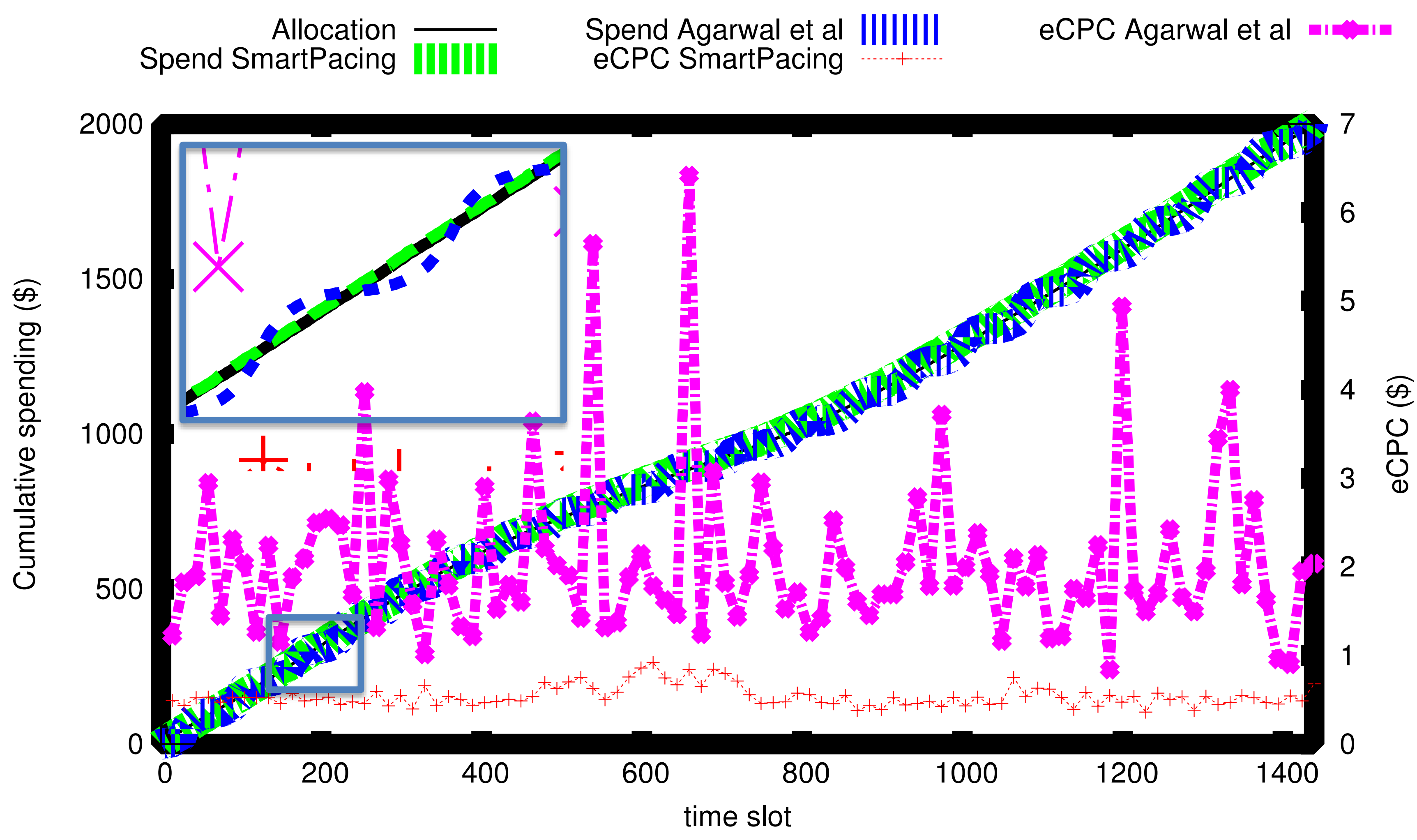
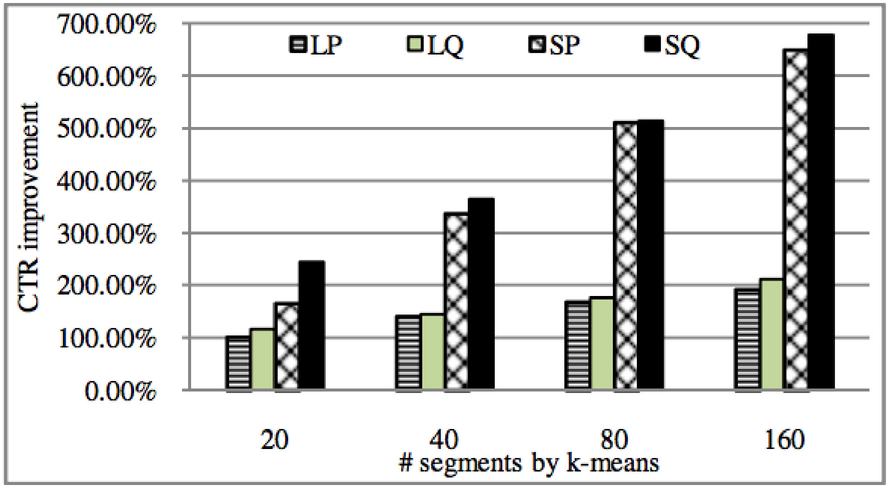


Table of contents

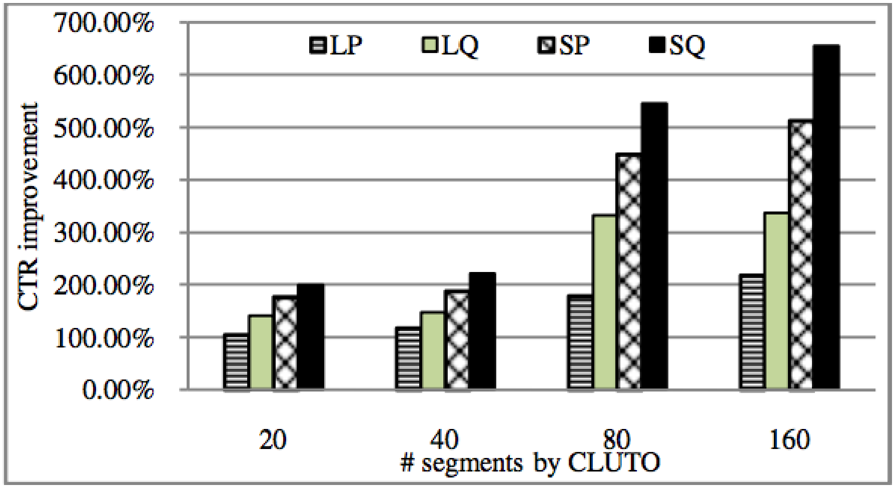
Does targeting help online advertising?

- Segment user based on …
  - LP: Long-term Page-view ,   SP: Short-term Page-view
  - LQ: Long-term Query ,   SQ: Short-term Query

Compare the best CTR segment with baseline (random users)



(a) User clustering by k-means

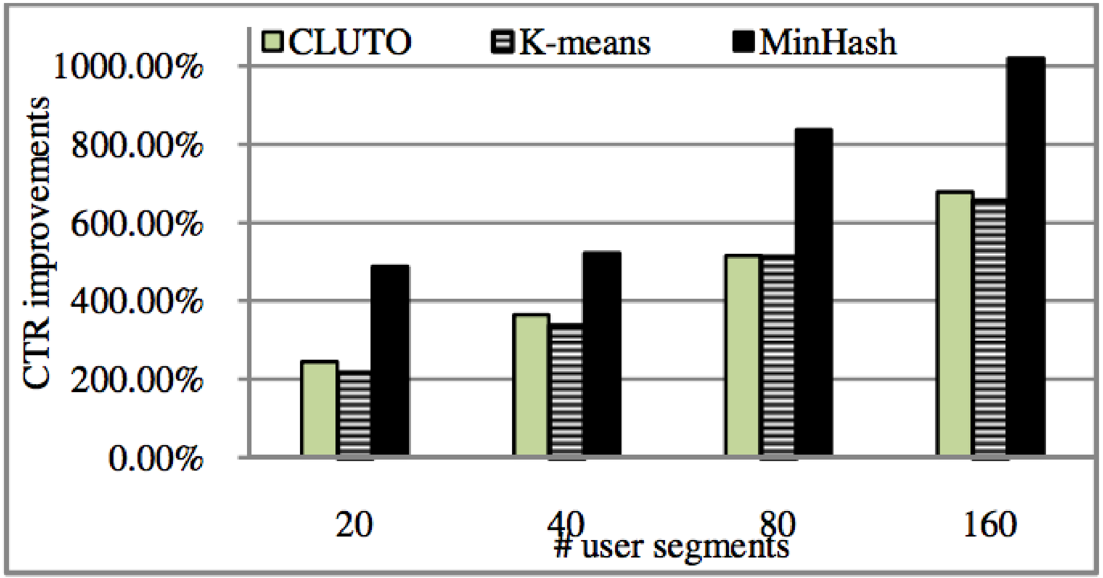


(b) User clustering by CLUTO

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

User segmentation

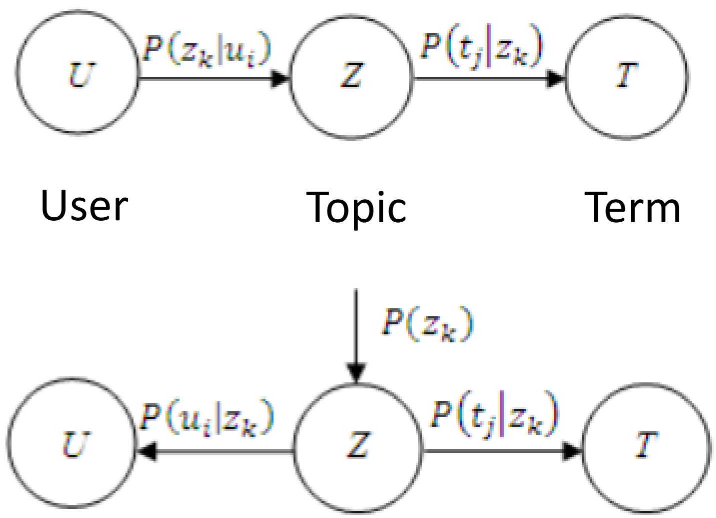
- Different user segmentation algorithms may have different results



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

User segmentation

- From user – documents to user – topics
  - Topic modeling using PLSA, LDA, etc.



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

Targeting landscape

- Targeting: reach the *precise* users who are receptive to the marketing messages.

Geo-targeting

Demo-targeting

Web-site targeting

Behavioral Targeting

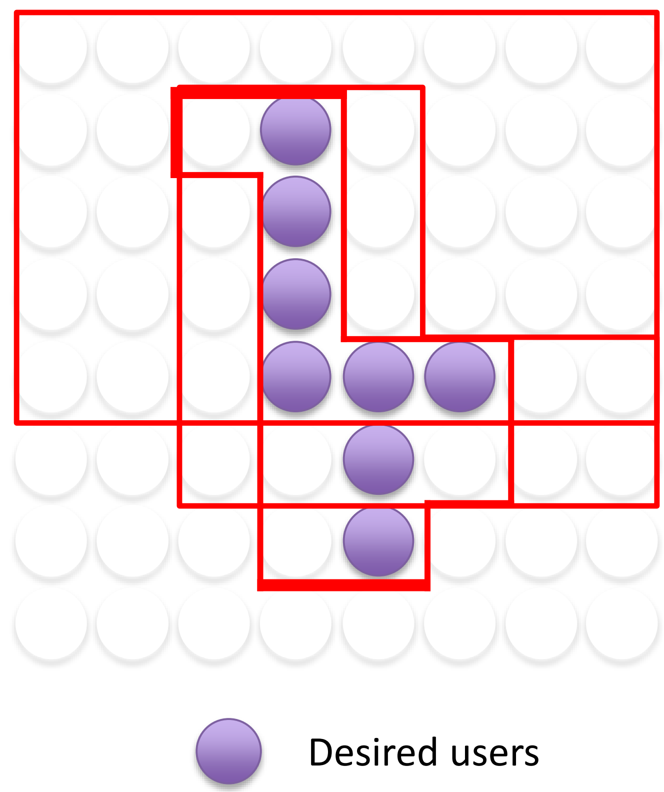
Site Re-targeting

Social Targeting

Search Re-targeting

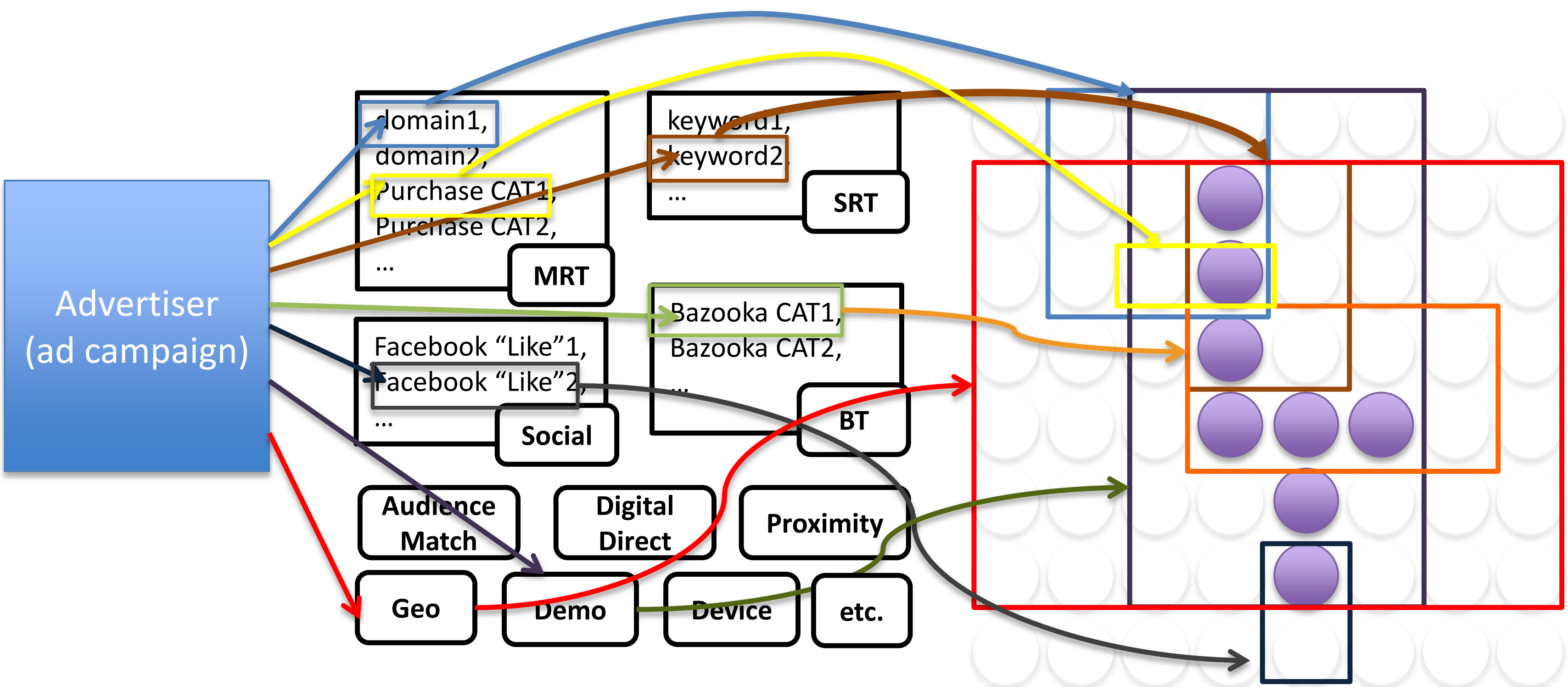
Mail Re-targeting

Proximity Targeting



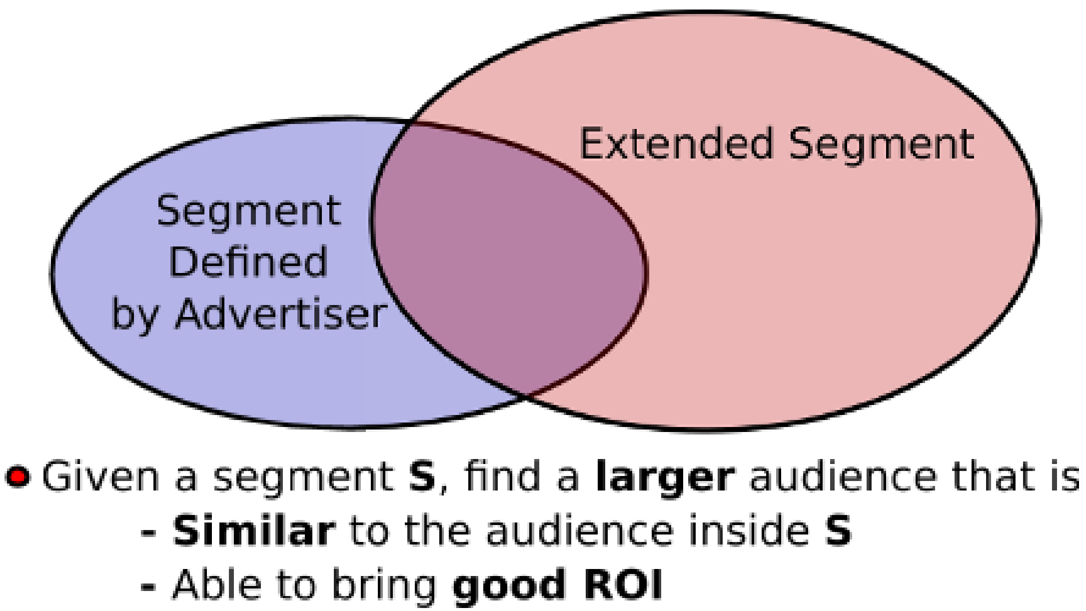
Targeting landscape

- A bit too complicated …



Audience expansion

- AEX Simplifies targeting by discovering similar (prospective) customers



[J Shen, et al., Effective Audience Extension in Online Advertising, KDD 2015]

Rule mining-based approach

- Identify feature-pair-based associative classification rules
  - Affinity that a *feature-pair* towards conversion:

$$F\text{-}LLR = \boxed{P(f)} \times log\left(\frac{P(f|conversion)}{P(f|non-conversion)}\right)$$

Probability to observe feature-pair *f* in data

  - Top *k* feature (pairs) are kept as scoring rules

Especially good for those tail campaigns (e.g. CVR < 0.01%)

[Mangalampalli et al, A feature-pair-based associative classification approach to look-alike modeling for conversion-oriented user-targeting in tail campaigns. WWW 2011]

Rule mining-based approach

- Campaign C1: a *tail* campaign
- Campaign C2: a *head* campaign

Table 1: Results for Campaign C_1

Baseline	Lift (Conversion Rate)	Lift (AUC)
Random Targeting	82%	–
Linear SVM	301%	11%
GBDT	100%	2%

Table 2: Results for Campaign C_2

Baseline	Lift (Conversion Rate)	Lift (AUC)
Random Targeting	48%	–
Linear SVM	-12%	-6%
GBDT	-40%	-14%

[Mangalampalli et al, A feature-pair-based associative classification approach to look-alike modeling for conversion-oriented user-targeting in tail campaigns. WWW 2011]

Weighted criteria-based approach

- Similarity Criterion:

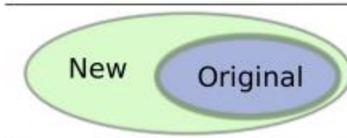
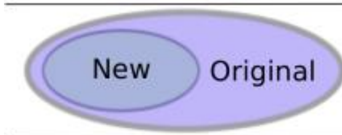
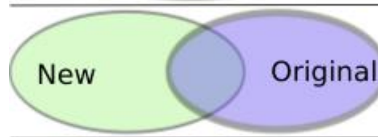
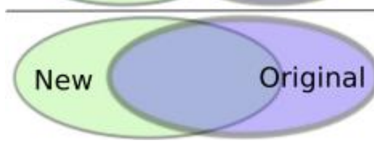
$$sim(c_{new}, S) = p(c_{new}|S)$$

$$= \frac{|aud(c_{new}) \cap aud(S)|}{|aud(S)|}$$

- Novelty Criterion:

$$nov(c_{new}, S) = p(!S|c_{new})$$

$$= 1 - p(S|c_{new})$$

	Similarity $P(New\|Original)$	Novelty $P(!Original\|New)$	Value $Good/OK/Bad?$
	1	0	Bad
	1	≈ 0.5	Good
	≈ 0.5	0	Bad
	≈ 0.2	≈ 0.8	OK
	≈ 0.8	≈ 0.2	OK

[J Shen, et al., Effective Audience Extension in Online Advertising, KDD 2015]

Weighted criteria-based approach

- Quality Criterion:

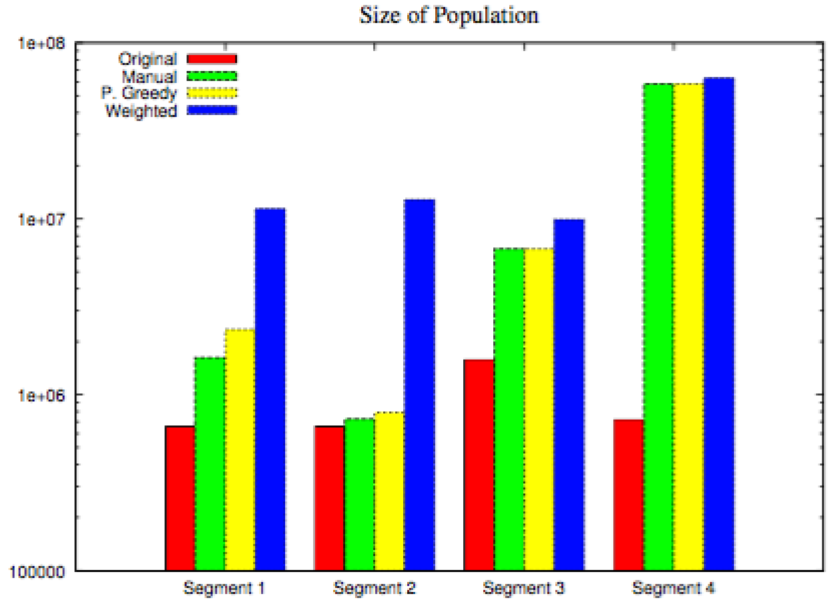
$$q(c_{new}) = \frac{\sum_{u \in aud(c_{new})} click(u, adv)}{\sum_{u \in aud(c_{new})} imp(u, adv)}$$

- Final score

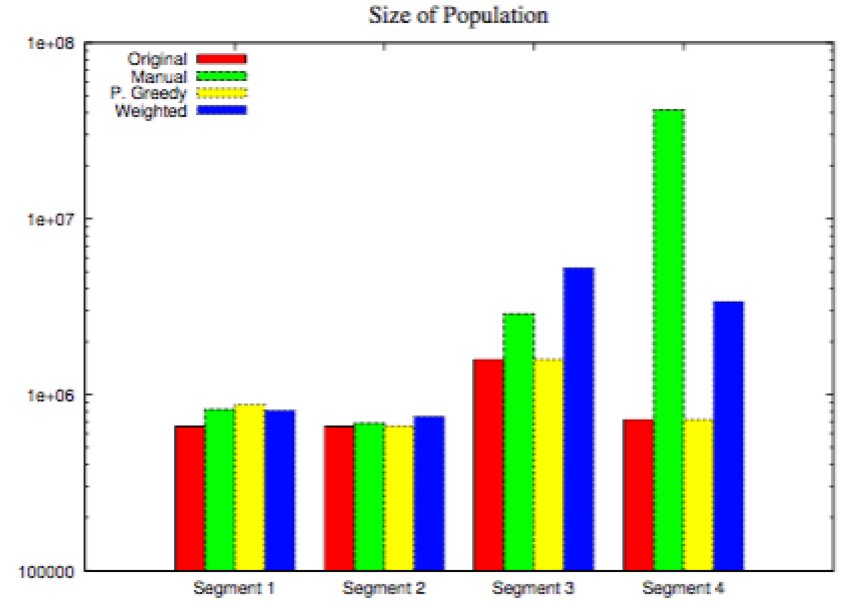
$$logScore(c_{new}|S) = \theta_1 \, log(p(c_{new}|S)) +$$

$$\theta_2 \, log(1 - p(S|c_{new})) \; + \; \theta_3 \, log(q(c_{new}))$$

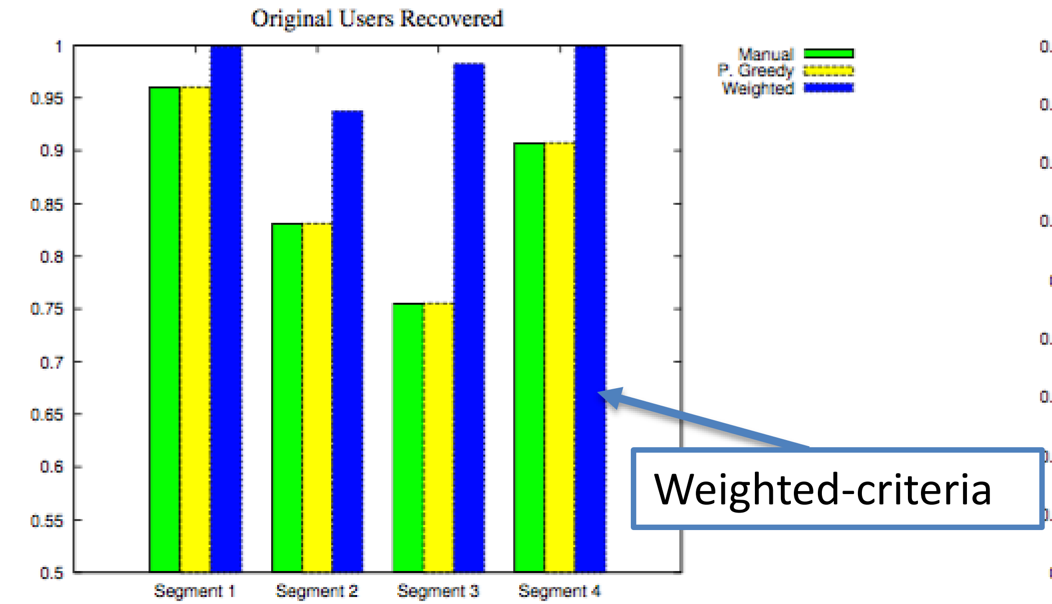
[J Shen, et al., Effective Audience Extension in Online Advertising, KDD 2015]



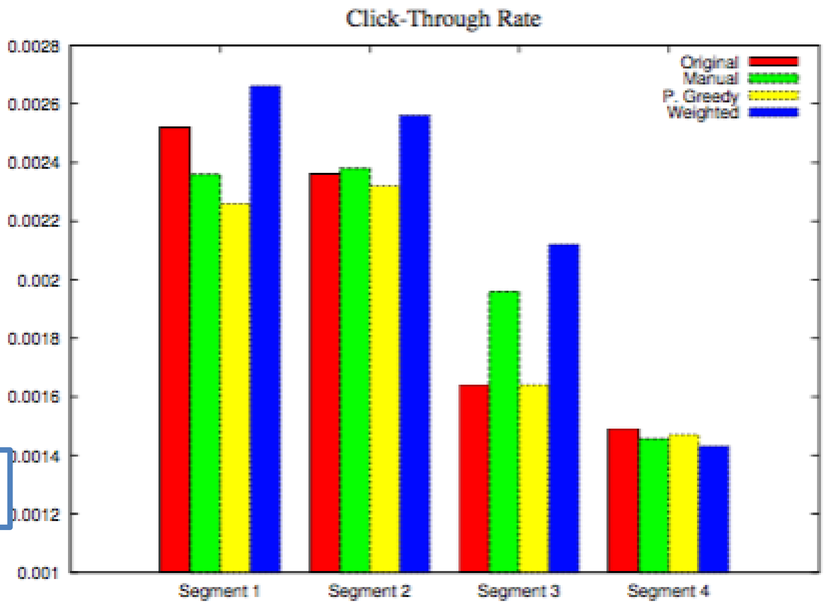
(a) Sizes of audiences for the original segment and different recommended extensions.



(a) Sizes of audiences for the original segment and different recommended extensions.



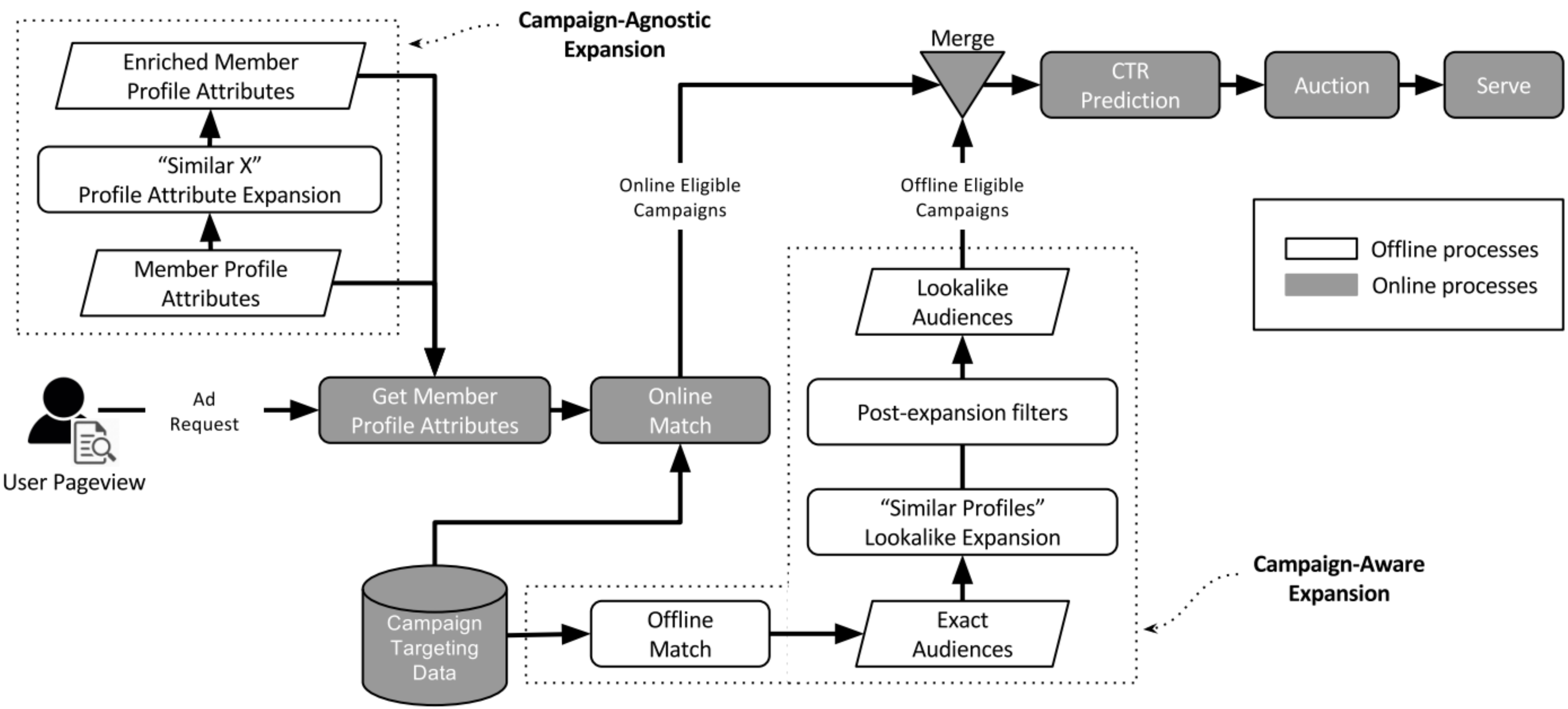
(b) Amount of original audience covered by different recommended extensions.



(b) CTR values for the original segment and different recommended extensions.

Audience Expansion for OSN Advertising

- Campaign-agnostic: enrich member profile attributes
- Campaign-aware: identify similar members



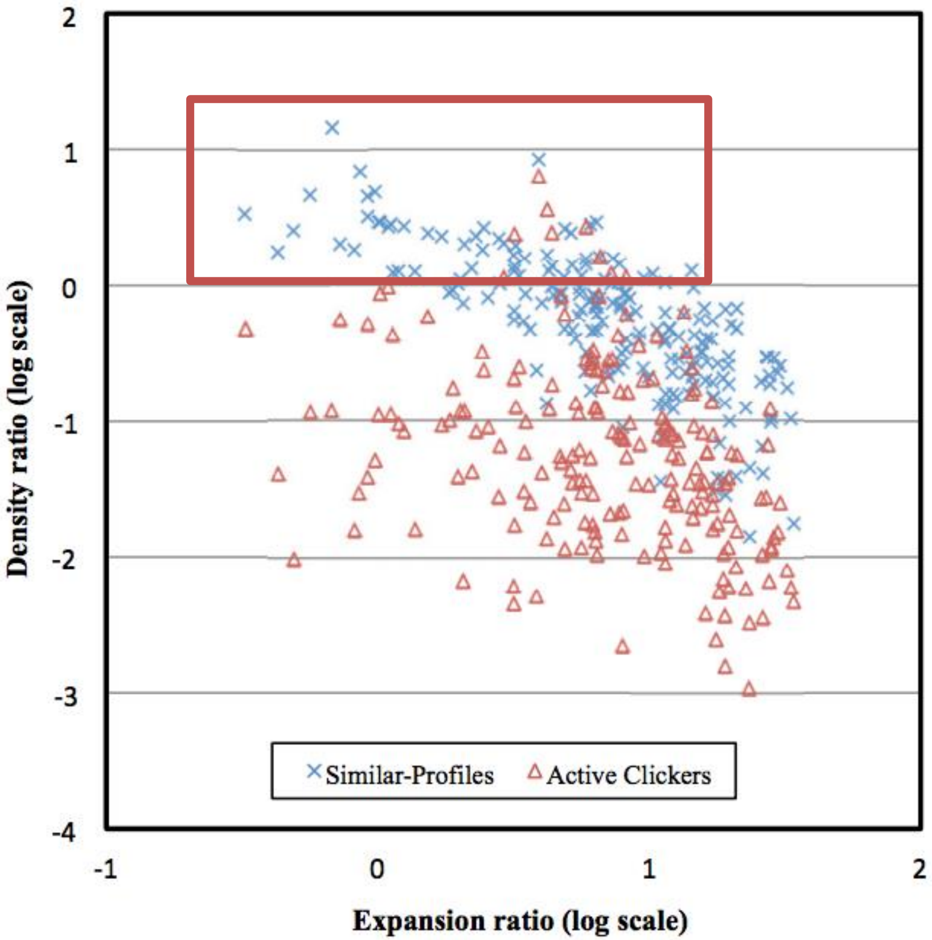
[H Liu et al. Audience expansion for online social network advertising. KDD 2016]

Audience Expansion for OSN Advertising

- Member similarity evaluation
  - Density of a segment:

$$D = \frac{2|C|}{|M|(|M|-1)}$$

  - Expansion ratio vs Density ratio



[H Liu et al. Audience expansion for online social network advertising. KDD 2016]

Transferred lookalike

- Web browsing prediction (CF task)

$$\hat{y}^{\mathrm{c}}_{u,p} = \sigma\Big(w^{\mathrm{c}}_0 + \sum_i w^{\mathrm{c}}_i \underbrace{\boxed{x^u_i}}_{\text{user feature}} + \sum_j w^{\mathrm{c}}_j \underbrace{\boxed{x^p_j}}_{\text{publisher feature}} + \sum_i \sum_j \langle \underbrace{\boxed{\boldsymbol{v}^{\mathrm{c}}_i}}_{\text{K-dimensional latent vector}}, \boldsymbol{v}^{\mathrm{c}}_j \rangle x^u_i x^p_j \Big)$$

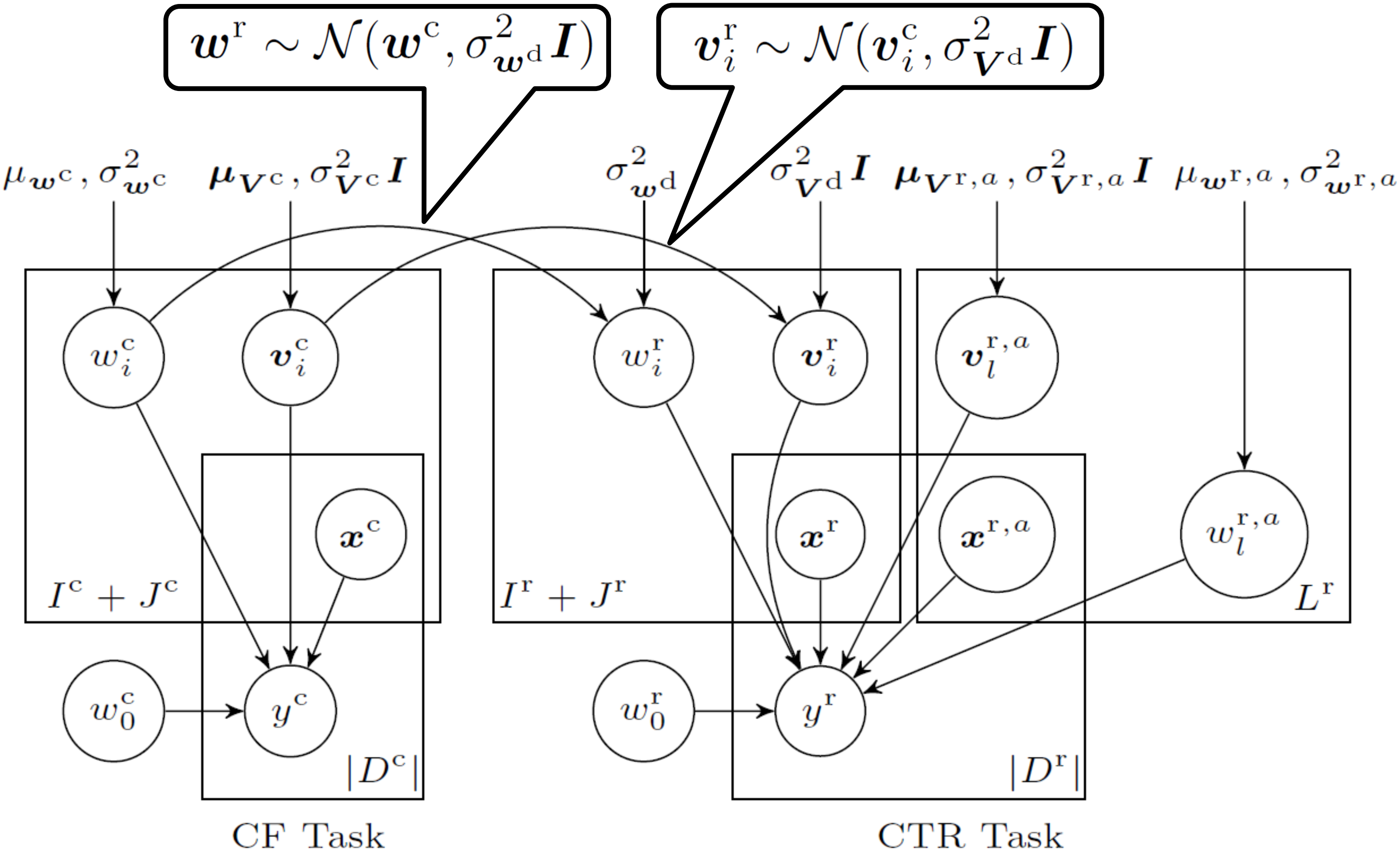
- Ad response prediction (CTR task)

$$\hat{y}^{\mathrm{r}}_{u,p,a} = \sigma\Big(w^{\mathrm{r}}_0 + \sum_i w^{\mathrm{r}}_i x^u_i + \sum_j w^{\mathrm{r}}_j x^p_j + \sum_l w^{\mathrm{r}}_l \underbrace{\boxed{x^a_l}}_{\text{ad feature}} + \sum_i \sum_j \langle \boldsymbol{v}^{\mathrm{r}}_i, \boldsymbol{v}^{\mathrm{r}}_j \rangle x^u_i x^p_j + \sum_i \sum_l \langle \boldsymbol{v}^{\mathrm{r}}_i, \boldsymbol{v}^{\mathrm{r}}_l \rangle x^u_i x^a_l + \sum_j \sum_l \langle \boldsymbol{v}^{\mathrm{r}}_j, \boldsymbol{v}^{\mathrm{r}}_l \rangle x^p_j x^a_l \Big)$$

[Zhang et al. Implicit Look-alike Modelling in Display Ads: Transfer Collaborative Filtering to CTR Estimation. ECIR 2016]

Transferred lookalike

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

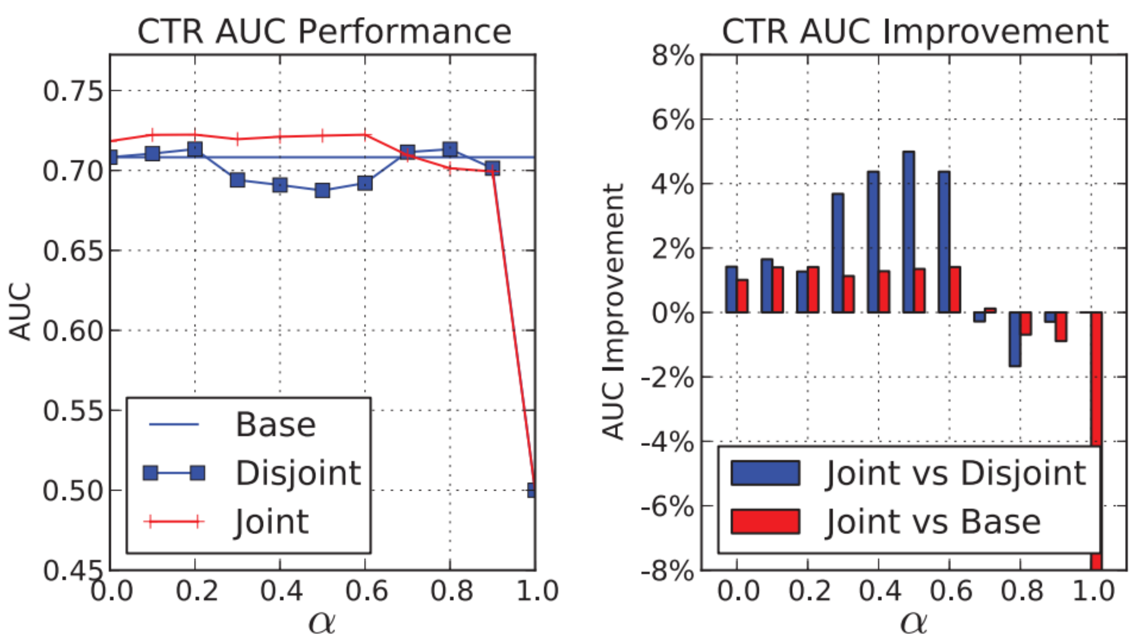


[Zhang et al. Implicit Look-alike Modelling in Display Ads: Transfer Collaborative Filtering to CTR Estimation. ECIR 2016]

Joint Learning in Transferred lookalike

$$\hat{\Theta} = \max_{\Theta} P(\Theta) \Big[\prod_{(\boldsymbol{x}^{c}, y^{c}) \in D^{c}} P(y^{c}|\boldsymbol{x}^{c};\Theta) \Big]^{\frac{\alpha}{|D^{c}|}} \cdot \Big[\prod_{(\boldsymbol{x}^{r}, y^{r}) \in D^{r}} P(y^{r}|\boldsymbol{x}^{r};\Theta) \Big]^{\frac{1-\alpha}{|D^{r}|}}$$

$$P(\Theta) = P(\boldsymbol{w}^{c}) P(\boldsymbol{V}^{c}) P(\boldsymbol{w}^{r}|\boldsymbol{w}^{c}) P(\boldsymbol{V}^{r}|\boldsymbol{V}^{c}) P(\boldsymbol{w}^{r,a}) P(\boldsymbol{V}^{r,a})$$

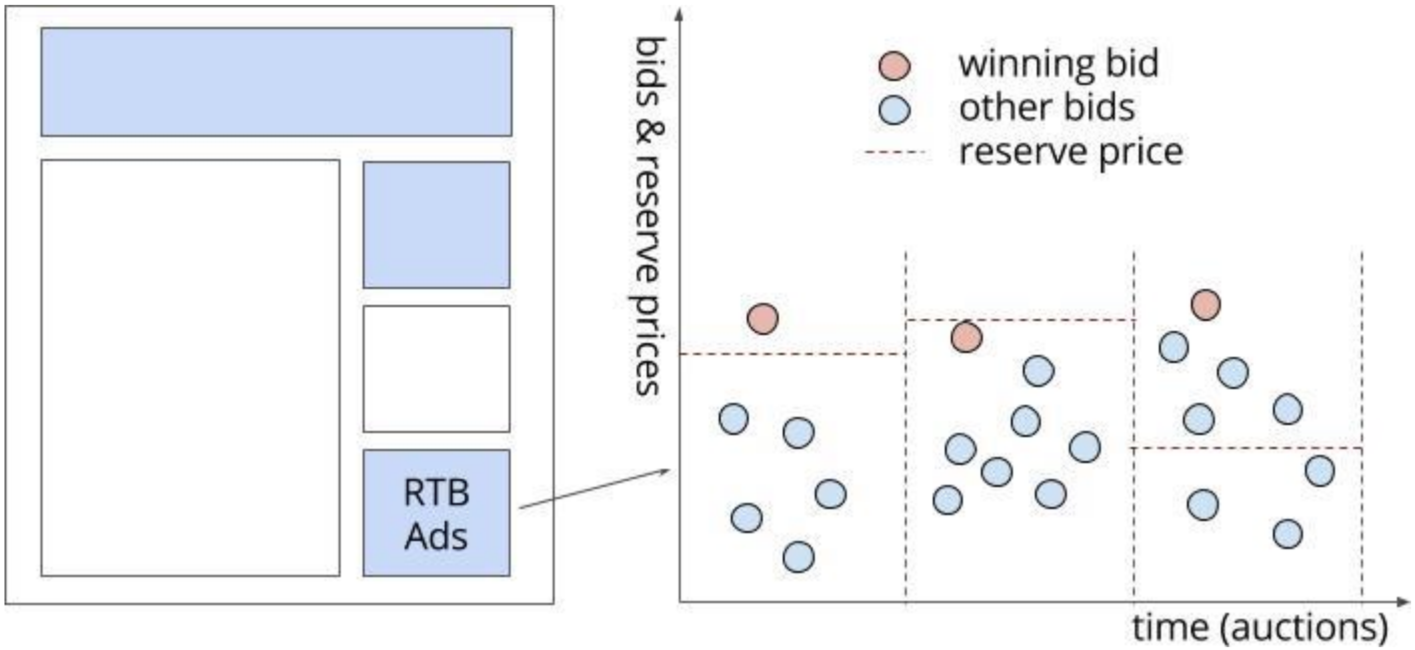


[Zhang et al. Implicit Look-alike Modelling in Display Ads: Transfer Collaborative Filtering to CTR Estimation. ECIR 2016]

Table of contents

- RTB system
- Auction mechanisms
- User response estimation
- Learning to bid
- Conversion attribution
- Pacing control
- Targeting and audience expansion
- Reserve price optimization

Reserve price optimisation



The task:

- To find the optimal reserve prices to maximize publisher revenue

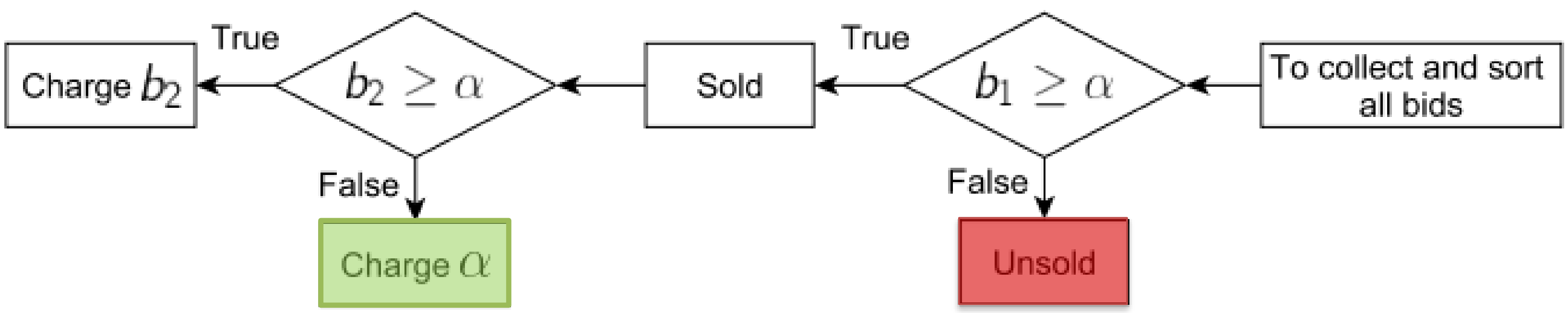
The challenge:

- Practical constraints v.s theoretical assumptions

[Yuan et al. An Empirical Study of Reserve Price Optimisation in Display Advertising. KDD 2014]

Why

- Suppose it is second price auction and b_1, b_2 are first and second prices
  - Preferable case: $b_1 \geq \alpha > \ b_2$ (increases revenue)
  - Undesirable case: $\alpha > b_1$ (lose revenue)



An example

- Suppose: two bidders, whose private values b_1, b_2 are both drawn from Uniform[0, 1]
- Without a reserve price, the expected payoff r is:

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

- With $\alpha = 0.2$:

$$r = E[\min(b_1, b_2)\,|b_1 > 0.2, b_2 > 0.2] + (0.8 \times 0.2) \times 2 \times 0.2 = 0.36$$

- With $\alpha = 0.5$:

$$r = E[\min(b_1, b_2)\,|b_1 > 0.5, b_2 > 0.5] + (0.5 \times 0.5) \times 2 \times 0.5 = 0.42$$

- With $\alpha = 0.6$:

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

Paying the second highest price

Paying the reserve price

[Ostrovsky et al, Reserve prices in internet advertising auctions: A field experiment. EC 2011]

Theoretically optimal reserve price

- In the second price auctions, an advertiser bid its private value b

- Suppose bidders are risk-neutral and symmetric (i.e. having same distributions) with bid C.D.F $F(b)$

- The publisher also has a private value V_p
- The optimal reserve price is given by: $\alpha = \dfrac{1 - F(\alpha)}{F'(\alpha)} + V_p$

[Levin and Smith, Optimal Reservation Prices in Auctions, 1996]

Results from a field experiment

- Using the theoretically optimal reserve price on Yahoo! Sponsored search

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

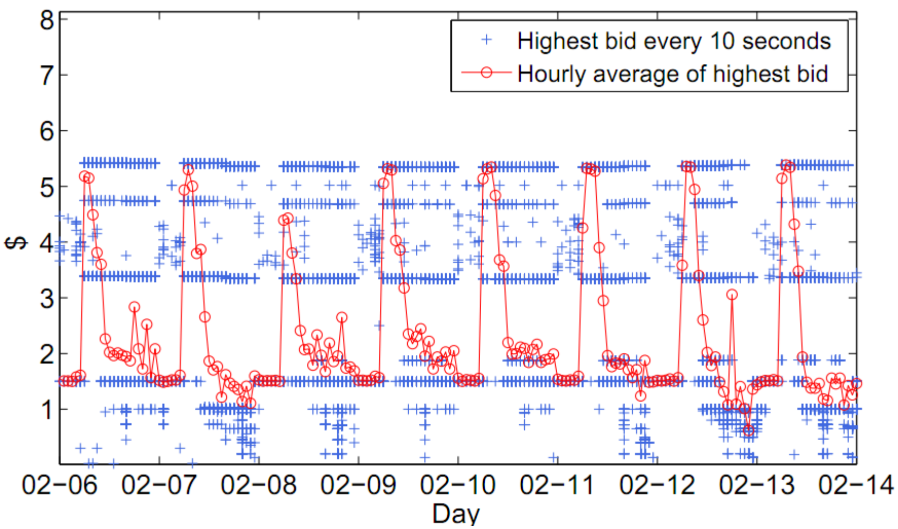
Table 8: Restricted sample (optimal reserve price $\geq 20¢$)

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

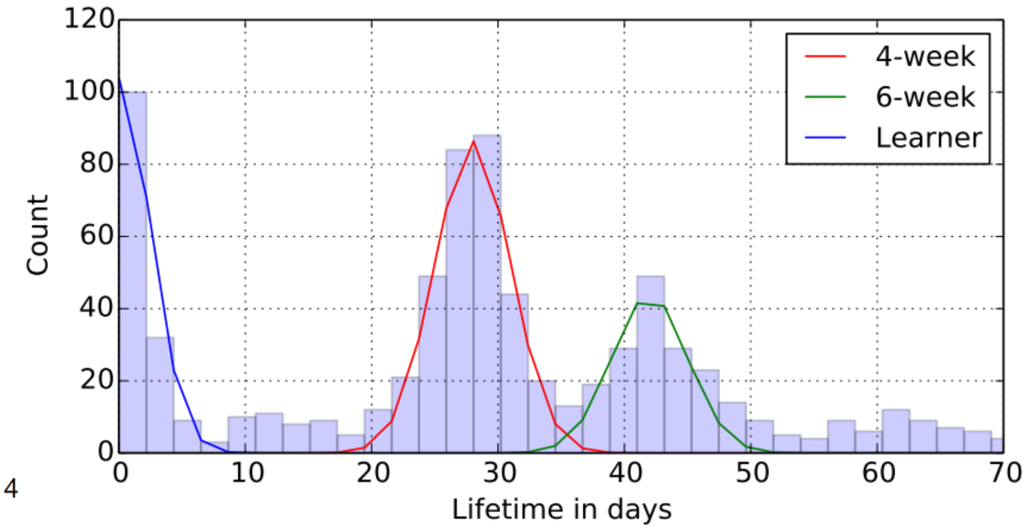
[Ostrovsky et al, Reserve prices in internet advertising auctions: A field experiment. EC 2011]

Bidding strategy is a mystery

- Advertisers have their own bidding strategies (No access to publishers)
- They change their strategies frequently



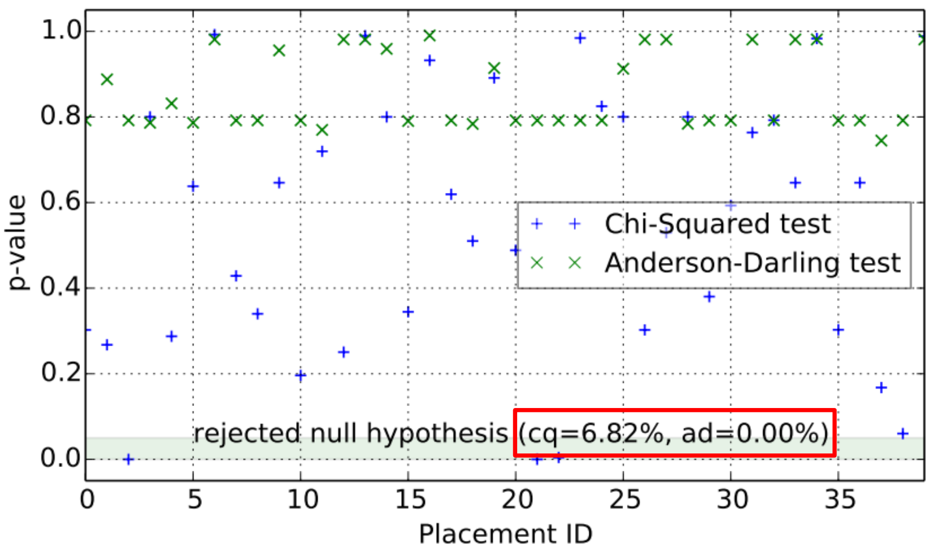
Many advertisers bid at fixed values with bursts and randomness.

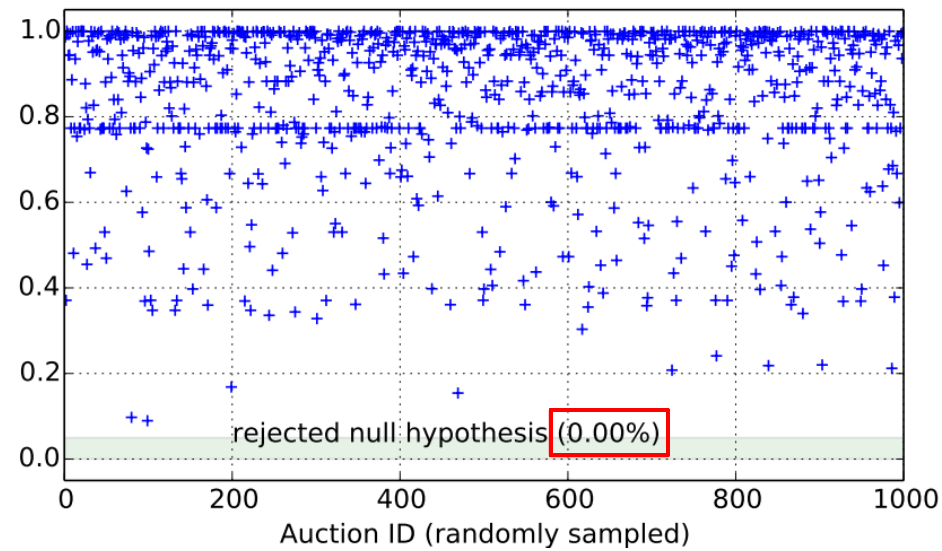


And they come and go

[Yuan et al. An Empirical Study of Reserve Price Optimisation in Display Advertising. KDD 2014]

Uniform/Log-normal distributions do NOT fit well





Test at the placement level
(because we usually set reserve prices on placements)

Test at the auction level

- Chi-squared test for Uniformity
- Anderson-Darling test for Normality

[Yuan et al. An Empirical Study of Reserve Price Optimisation in Display Advertising. KDD 2014]

A simplified dynamic game

- Players: auction winner w ,publisher p
- Initial status: $I_1 : b \geq \alpha$; I_2 otherwise

- The action set of the winner A_w:
  a_{w1}, to increase b to higher than α;
  a_{w2}, to increase b to lower than α;
  a_{w3}, to decrease or hold b to higher than α;
  a_{w4}, to decrease or hold b to lower than α.
- The action set of the publisher A_p:
  a_{p1}, to increase or hold α to higher than b;
  a_{p2}, to increase or hold α to lower than b;
  a_{p3}, to decrease α to higher than b;
  a_{p4}, to decrease α to lower than b.

$$s_p^*(I) = \begin{cases} a_{p2}, & \text{if } I = I_1 \\ a_{p4}, & \text{if } I = I_2 \end{cases}$$

$$s_w^*(I) = \begin{cases} a_{w3}, & \text{if } I = I_1 \\ a_{w1}, & \text{if } I = I_2 \end{cases}$$

[Yuan et al. An Empirical Study of Reserve Price Optimisation in Display Advertising. KDD 2014]

OneShot: the algorithm based on dominant strategy

- The algorithm essentially uses a conventional feedback controller

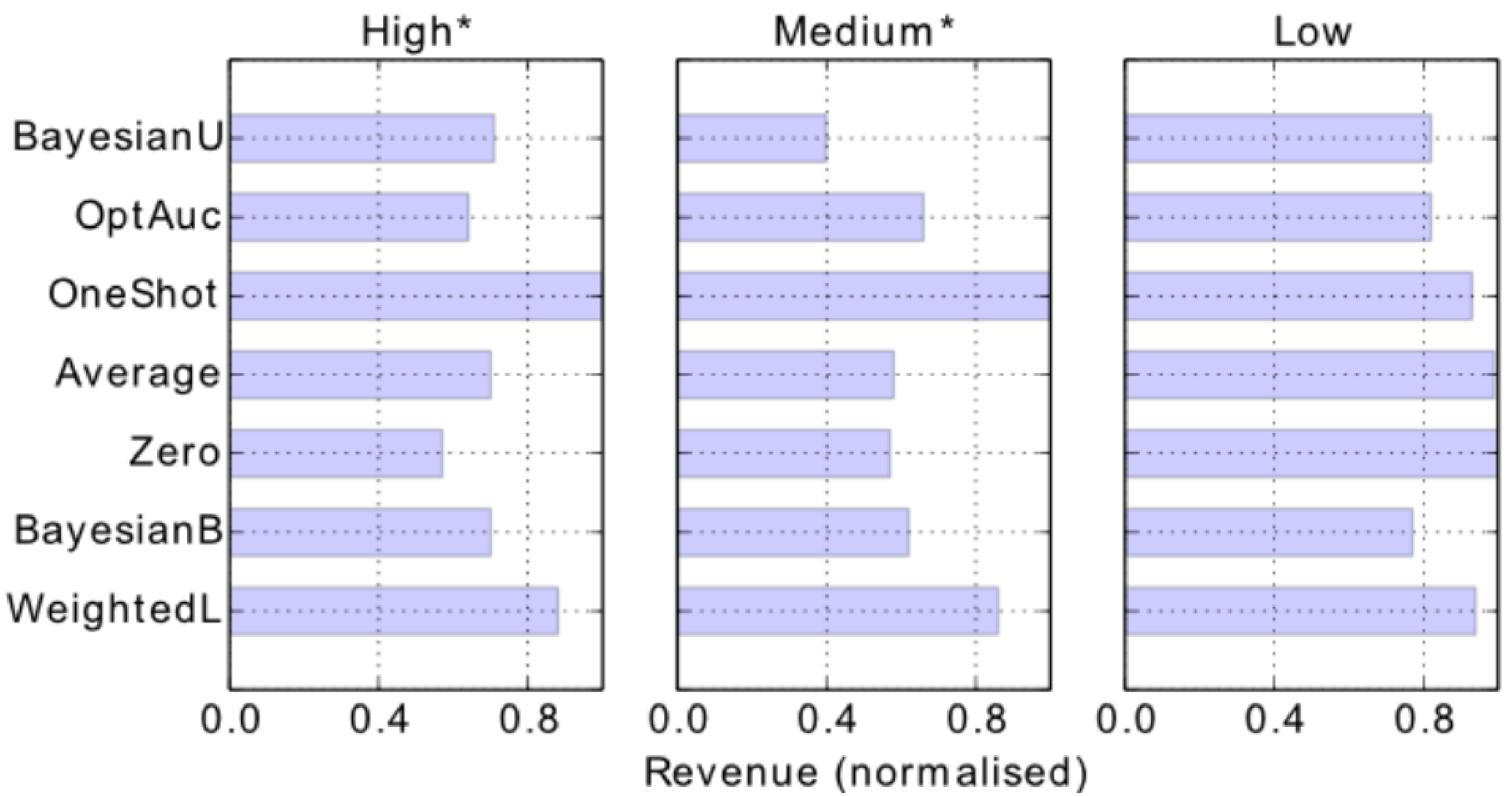
$$
\begin{cases}
\alpha(t+1) = (1 - \epsilon^t \lambda_h) a(t) & \text{if } \alpha(t) > b_1(t) \\
\alpha(t+1) = (1 + \epsilon^t \lambda_e) a(t) & \text{if } b_1(t) \geq \alpha(t) \geq b_2(t) \\
\alpha(t+1) = (1 + \epsilon^t \lambda_l) \alpha(t) & \text{if } b_2(t) > \alpha(t)
\end{cases}
$$

- A practical example setting of the parameters:

$$\epsilon = 1.0,\ \lambda_h = 0.3,\ \lambda_e = 0.01, \text{ and } \lambda_l = 0.02$$

[Yuan et al. An Empirical Study of Reserve Price Optimisation in Display Advertising. KDD 2014]

OneShot performance



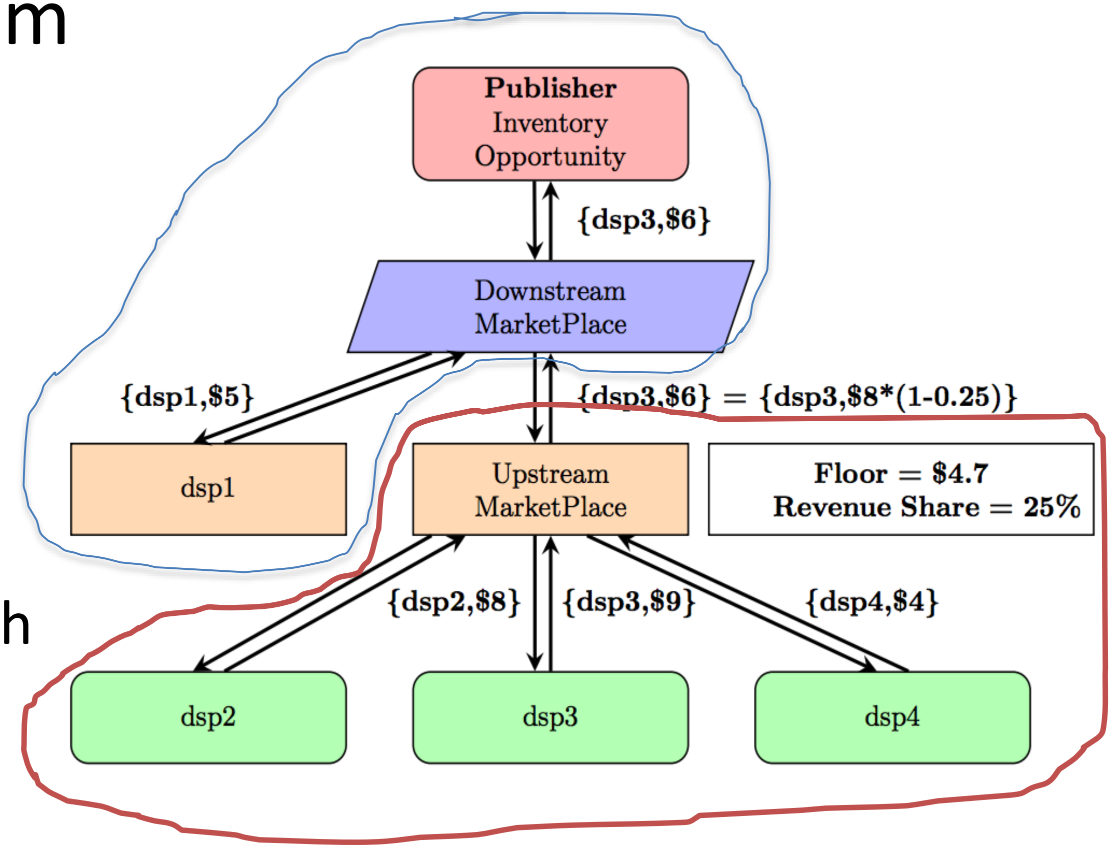
[Yuan et al. An Empirical Study of Reserve Price Optimisation in Display Advertising. KDD 2014]

Advertiser attrition concern

[Yuan et al. An Empirical Study of Reserve Price Optimisation in Display Advertising. KDD 2014]

Optimal reserve price in upstream auctions

- A different problem setting
  - Upstream charges a revenue-share (e.g. 25%) from each winning bid.
  - What is the optimal reserve price for such a marketplace?



[Alcobendas et al., Optimal reserve price in upstream auctions: Empirical application on online video advertising. KDD 2016]

Optimal reserve price in upstream auctions

- Assume bidder's valuation of the inventory is an i.i.d. realization of the random variable V, and bidders are risk neutral, the optimal reserve price for upstream marketplace satisfies

Probability of winning downstream auction

$$[\rho_c^* f_V(\rho_c^*) - 1 + F_V(\rho_c^*)]\, F_V(\rho_c^*)^{\mathcal{N}-1} \boxed{P_D(\rho_c^*)} =$$

Support interval of V

$$\frac{\partial P_D(\boxed{\bar{w}_T(\rho_c^*)})}{\partial \rho_c^*} \int_{\rho_c^*}^{\boxed{\bar{v}}} [u f_V(u) - 1 + F(u)]\, \boxed{F_V(u)^{\mathcal{N}-1}} du$$

Expected price if having at least one bidder above reserve price

Probability that a bidder wins the upstream auction with bid u

If without downstream auction, optimal condition is

$$[\rho_u^* f_V(\rho_u^*) - 1 + F_V(\rho_u^*)] = 0$$

Optimal reserve price in upstream auctions

Type of Placement	Nb Placements	Placements with Positive Revenue Lift (%)	Expected Revenue Lift (%)
No Downstream Auction (ρ_u^*)	71	77%	39%
Downstream Auction: No Correction (ρ_u^*)	30	67%	$\boxed{25\%}$
Downstream Auction: Correction (ρ_c^*)	30	77%	$\boxed{29\%}$

Type of Placement	Nb Placements	Placements with Positive Revenue Lift (%)	Expected Revenue Lift (%)
No Downstream Auction (ρ_u^*)			
- Above Current Floor	24	88%	38%
- Below Current Floor	47	72%	40%
Downstream Auction: No Correction (ρ_u^*)			
- Above Current Floor	9	100%	92%
- Below Current Floor	21	52%	11%
Downstream Auction: Correction (ρ_c^*)			
- Above Current Floor	13	100%	88%
- Below Current Floor	17	71%	22%

[Alcobendas et al., Optimal reserve price in upstream auctions: Empirical application on online video advertising. KDD 2016]

CIKM16 Tutorial

Learning, Prediction and Optimisation in RTB Display Advertising

Thank You

- RTB system
- Auction mechanisms
- User response estimation
- Learning to bid
- Conversion attribution
- Pacing control
- Targeting and audience expansion
- Reserve price optimization

Weinan Zhang
(wnzhang AT sjtu.edu.cn)

Jian Xu
(jian.xu AT cootek.cn)