User Response Learning for Directly Optimizing Campaign Performance in Display Advertising

Kan Ren[1] Weinan Zhang[1] Yifei Rong[3]
Haifeng Zhang[4] Yong Yu[1] Jun Wang[2]

[1]Apex Data and Knowledge Management Lab
Shanghai Jiao Tong University

[2]University College London & MediaGamma Ltd

[3]YOYI Inc.

[4]Peking University

CIKM 2016

Outline

1. User Response Prediction
   - Goal and Applications
   - State-of-the-art Methods
2. Case Study: Real-time Bidding
3. Optimize the Response Prediction
   - Problem Setup
   - Optimization for Campaign Performance
     - Expected Utility Model
     - Risk Return Model
   - Realization of Model
   - Links to Previous Work
4. Experiments
   - Evaluations

Outline

User Response Prediction

Goal

To estimate the *probability of user taking some specific actions.*

Example

Click-through Rate (CTR): Possibility of user clicking on the specific item.
Conversion Rate (CVR): Possibility of user convert an action on the item.

Outline

State-of-the-art

Regression Model

Logistic Regression (LR)

Estimating Conversion Rate in Display Advertising from Past Performance Data, K.-c Lee et al. KDD 2012

Tree-based Model

Practical Lessons from Predicting Clicks on Ads at Facebook, X. He et al. ADKDD 2014

Factorization Machines

Response prediction using collaborative filtering with hierarchies and side-information, A.K. Menon et al. KDD 2011

State-of-the-art

Other Variants

Bayesian Probit Regression

Web-scale Bayesian Click-through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine, T. Graepel et al. ICML 2010

Factorization Machine with FTRL

Factorization Machines with Follow-The-Regularized-Leader for CTR prediction in Display Advertising, A.-P. Ta. Big Data 2015

Deep Neural Networks

A Convolutional Click Prediction Model, Q. Liu et al. CIKM 2015
Deep Learning over Multi-field Categorical Data A Case Study on User Response Prediction. W. Zhang et al. ECIR 2016.

Objective Function

Squared Error

$$\mathcal{L}_{SE} = \frac{1}{2}(y - \hat{y})^2, y \in \{0, 1\}, \hat{y} \in [0, 1]$$

Cross Entropy

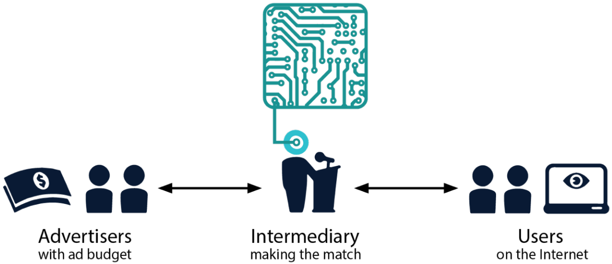
$$\mathcal{L}_{CE} = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}), y \in \{0, 1\}, \hat{y} \in [0, 1]$$

Evaluation Measurements

Area under ROC Curve (AUC)

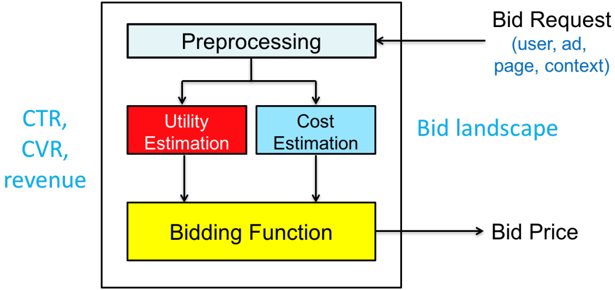
Relative Information Gain (Cross Entropy)

RTB Display Advertising



To address the **right user** *with the* **right message** *in the* **right context** *and at the* **right price**.

Bidding Strategy



Advertisers try to

- buy ad impression opportunities
- pursuing best returns (profit = gain - cost)

Problems & Challenges

Isolated Optimization

Models for different components are trained separately.

Blind of Sequential Stages

Impression opportunities of low price with high value and high price with low value have not been appropriately dealt with.

Incomplete Evaluations [1]

Common metrics cannot reflect the true concern of the advertisers. (high AUC does not correspond to high profit)

[1] Offline Evaluation of Response Prediction in Online Advertising Auctions, O. Chapelle. WWW 2015

Outline

1. User Response Prediction
   - Goal and Applications
   - State-of-the-art Methods
2. Case Study: Real-time Bidding
3. Optimize the Response Prediction
   - Problem Setup
   - Optimization for Campaign Performance
     - Expected Utility Model
     - Risk Return Model
   - Realization of Model
   - Links to Previous Work
4. Experiments
   - Evaluations

Problem Setup

Motivation

To make the user response prediction more precisely accurate considering the context.

Problem Setup

Motivation

To make the user response prediction more precisely accurate considering the context.

Example

Combine response learning and bidding strategy, market price modeling altogether.
That is to directly optimize profit for a performance ad campaign, rather than only optimizing discrete user responses.

Outline

Notations and descriptions

Notation	Description
y	The true label of user response.
$\boldsymbol{x}$	The bid request represented by its features.
$\boldsymbol{\theta}$	The parameter of CTR estimation function.
$f_{\boldsymbol{\theta}}(\boldsymbol{x})$	the CTR estimation function to learn.
$b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))$	The bid price determined by the estimated CTR, b for short.
$R_{\boldsymbol{\theta}}(\cdot)$	The utility function.

Market Modeling

Market Price

The second highest price proposed during an RTB auction (2^{nd} price).

Market Price Distribution (p.d.f.)

$$p_z(z), z \in \mathcal{N}.$$

Market Modeling

Market Price

The second highest price proposed during an RTB auction (2^{nd} price).

Market Price Distribution (p.d.f.)

$$p_z(z), z \in \mathcal{N}.$$

Winning Probability when Bidding at price b (c.d.f.)

$$w(b) = \int_0^b p_z(z)dz. \tag{1}$$

Market Modeling

Market Price

The second highest price proposed during an RTB auction (2^{nd} price).

Market Price Distribution (p.d.f.)

$$p_z(z), z \in \mathcal{N}.$$

Winning Probability when Bidding at price b (c.d.f.)

$$w(b) = \int_0^b p_z(z)dz. \tag{1}$$

Expected Cost under 2^{nd} Price Auction

$$\int_0^b zp_z(z)dz \qquad \int_0^b zp_z(z)dz$$

Objective Function

Objective Function

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \int_{\boldsymbol{x}} R_{\boldsymbol{\theta}}(\boldsymbol{x}, y; b, v, c, w) p_{\boldsymbol{x}}(\boldsymbol{x}) d\boldsymbol{x}. \tag{3}$$

$R_{\boldsymbol{\theta}}(\cdot)$ is the *Utility Function.*

Constant click value v limits the max bid.

We will propose two variants of $R_{\boldsymbol{\theta}}(\cdot)$.

Expected Utility Model

Expected Utility (EU)

$$R_{\boldsymbol{\theta}}^{\mathrm{EU}}(\boldsymbol{x}, y) = [vy - c(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))] \cdot w(b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))). \tag{4}$$

Objective of EU

The overall expected direct profit[2] of all the auctions can be calculated by replacing Eqs. (1) and (2) into Eq. (4) as

$$\sum_{(\boldsymbol{x},y)\in D} R_{\boldsymbol{\theta}}^{\mathrm{EU}}(\boldsymbol{x},y) = \sum_{(\boldsymbol{x},y)\in D} [vy - c(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))] \cdot w(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))$$
$$= \sum_{(\boldsymbol{x},y)\in D} \left[vy - \frac{\int_0^{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} z \cdot p_z(z) dz}{\int_0^{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} p_z(z) dz} \right] \cdot \int_0^{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} p_z(z) dz$$
$$= \sum_{(\boldsymbol{x},y)\in D} \int_0^{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} (vy - z) \cdot p_z(z) dz. \quad (5)$$

[2] **O. Chapelle. WWW 2015**.

Optimal Parameter

Taking Eq. (5) into Eq. (3) with a regularization term turns our learning problem into convex optimization:

$$\begin{aligned}\boldsymbol{\theta}^{\mathrm{EU}} &= \arg\min_{\boldsymbol{\theta}} \; -\sum_{(\boldsymbol{x},y)\in D} R^{\mathrm{EU}}_{\boldsymbol{\theta}}(\boldsymbol{x},y) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2 \\ &= \arg\min_{\boldsymbol{\theta}} \; \sum_{\boldsymbol{x}} \int_0^{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} (z - vy)\cdot p_z(z)dz + \frac{\lambda}{2}\boldsymbol{\theta}^T\boldsymbol{\theta}.\end{aligned} \tag{6}$$

where the optimal value of $\boldsymbol{\theta}$ is obtained by taking a gradient descent algorithm.

Gradient of EU

The gradient of $R_{\boldsymbol{\theta}}^{\text{EU}}(\boldsymbol{x}, y)$ with regard to $\boldsymbol{\theta}$ is calculated as

$$\frac{\partial R_{\boldsymbol{\theta}}^{\text{EU}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = \overbrace{\big(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})) - vy\big)}^{\text{bid error}} \cdot \overbrace{p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))}^{\text{market sensitivity}} \cdot \frac{\partial b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})} \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta}. \tag{7}$$

and we update for each data instance as $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \frac{\partial R_{\boldsymbol{\theta}}^{\text{EU}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}}$ by above chain rule. (SGD)

Risk Return Model

Risk Return (RR)

$$R_{\boldsymbol{\theta}}^{\mathrm{RR}}(\boldsymbol{x}, y) = \Big(\underbrace{\frac{vy}{z}}_{\text{return}} - \underbrace{\frac{v(1-y)}{v-z}}_{\text{risk}} \Big) \cdot w(b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))). \tag{8}$$

Objective of RR

$$
\begin{aligned}
\boldsymbol{\theta}^{\mathrm{RR}} &= \arg\min_{\boldsymbol{\theta}} \; - \sum_{(\boldsymbol{x},y)\in D} R^{\mathrm{RR}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2 \\
&= \arg\min_{\boldsymbol{\theta}} \; - \sum_{(\boldsymbol{x},y)\in D} \int_0^{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} \left(\frac{vy}{z} - \frac{v(1-y)}{v-z}\right) \cdot p_z(z)dz \\
&\quad + \frac{\lambda}{2}\boldsymbol{\theta}^T\boldsymbol{\theta}. \qquad (9)
\end{aligned}
$$

Gradient of RR

$$\frac{\partial R_{\boldsymbol{\theta}}^{\text{RR}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = \Big(\overbrace{-\frac{vy}{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} + \frac{v(1-y)}{v - b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}}^{\text{bid error}}\Big) \cdot \overbrace{p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))}^{\text{market sensitivity}} \cdot \frac{\partial b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})} \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta}. \tag{10}$$

Outline

Question

Gradient of EU

$$\frac{\partial R^{\text{EU}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = \overbrace{(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})) - vy)}^{\text{bid error}} \cdot \overbrace{p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))}^{\text{market sensitivity}} \cdot \frac{\partial b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})} \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta}.$$

Gradient of RR

$$\frac{\partial R^{\text{RR}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = \Big(\overbrace{-\frac{vy}{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} + \frac{v(1-y)}{v - b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}}^{\text{bid error}}\Big) \cdot \overbrace{p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))}^{\text{market sensitivity}} \cdot \frac{\partial b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})} \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta}.$$

Question

Gradient of EU

$$\frac{\partial R^{\text{EU}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = \overbrace{(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})) - vy)}^{\text{bid error}} \cdot \overbrace{p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))}^{\text{market sensitivity}} \cdot \frac{\partial b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})} \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta}.$$

Gradient of RR

$$\frac{\partial R^{\text{RR}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = \Big(\overbrace{-\frac{vy}{b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))} + \frac{v(1-y)}{v - b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}}^{\text{bid error}}\Big) \cdot \overbrace{p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))}^{\text{market sensitivity}} \cdot \frac{\partial b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))}{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})} \frac{\partial f_{\boldsymbol{\theta}}(\boldsymbol{x})}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta}.$$

Realization of Model

Response Prediction Model $f_{\boldsymbol{\theta}}(\boldsymbol{x})$

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) \equiv \sigma(\boldsymbol{\theta}^T \boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}}} \ . \tag{11}$$

Linear Bidding Strategy

$$b(f_{\boldsymbol{\theta}}(\boldsymbol{x})) \equiv \phi \cdot v \cdot f_{\boldsymbol{\theta}}(\boldsymbol{x}), \tag{12}$$

where ϕ is the scaling parameter.

Linear Gradient

Linear Gradient of EU

$$\begin{aligned}\frac{\partial R^{\mathrm{EU}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} =& \phi v^2(\phi\sigma(\boldsymbol{\theta}^T\boldsymbol{x}) - y) \cdot p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))) \cdot \\ & \sigma(\boldsymbol{\theta}^T\boldsymbol{x})(1-\sigma(\boldsymbol{\theta}^T\boldsymbol{x}))\boldsymbol{x} + \lambda\boldsymbol{\theta}.\end{aligned} \tag{13}$$

Linear Gradient of RR

$$\begin{aligned}\frac{\partial R^{\mathrm{RR}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} =& \phi v\Big(-\frac{y}{\phi\sigma(\boldsymbol{\theta}^T\boldsymbol{x})} + \frac{1-y}{1-\phi\sigma(\boldsymbol{\theta}^T\boldsymbol{x})}\Big) \cdot \\ & p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))) \cdot \sigma(\boldsymbol{\theta}^T\boldsymbol{x})(1-\sigma(\boldsymbol{\theta}^T\boldsymbol{x}))\boldsymbol{x} + \lambda\boldsymbol{\theta}.\end{aligned} \tag{14}$$

Outline

Traditional Logistic Regression

Squared Error LR (SE)

$$
\begin{aligned}
\mathcal{L}^{\text{SE}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y) &= \frac{1}{2}(y - \sigma(\boldsymbol{\theta}^T \boldsymbol{x}))^2, \\
\frac{\partial \mathcal{L}^{\text{SE}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} &= (\sigma(\boldsymbol{\theta}^T \boldsymbol{x}) - y)\sigma(\boldsymbol{\theta}^T \boldsymbol{x})(1 - \sigma(\boldsymbol{\theta}^T \boldsymbol{x}))\boldsymbol{x}.
\end{aligned} \tag{15}
$$

Cross Entropy LR (CE)

$$
\begin{aligned}
\mathcal{L}^{\text{CE}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y) &= -y \log \sigma(\boldsymbol{\theta}^T \boldsymbol{x}) - (1 - y)\log(1 - \sigma(\boldsymbol{\theta}^T \boldsymbol{x})), \\
\frac{\partial \mathcal{L}^{\text{CE}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} &= (\sigma(\boldsymbol{\theta}^T \boldsymbol{x}) - y)\boldsymbol{x}.
\end{aligned} \tag{16}
$$

Discussion: Truthful Bidding Simplification

Simplification: Truthful Bidding, $\phi = 1$

$$b(f_{\boldsymbol{\theta}}(\boldsymbol{x})) = v \cdot f_{\boldsymbol{\theta}}(\boldsymbol{x}). \tag{17}$$

Corresponding Gradient of EU & RR

$$\begin{aligned}\frac{\partial R^{\text{EU}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} &= v^2(\sigma(\boldsymbol{\theta}^T\boldsymbol{x}) - y) \cdot p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x}))) \\ &\quad \cdot \sigma(\boldsymbol{\theta}^T\boldsymbol{x})(1 - \sigma(\boldsymbol{\theta}^T\boldsymbol{x}))\boldsymbol{x} + \lambda\boldsymbol{\theta}, \quad (18)\\ \frac{\partial R^{\text{RR}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} &= v(\sigma(\boldsymbol{\theta}^T\boldsymbol{x}) - y)p_z(b(f_{\boldsymbol{\theta}}(\boldsymbol{x})))\boldsymbol{x} + \lambda\boldsymbol{\theta}, \quad (19)\end{aligned}$$

EU & RR have one more component (market sensitivity) than SE & CE.

Discussion: Uniform Market Price Distribution Simplification

Simplification: Uniform Market Price Distribution

$$p_z(z) = l. \tag{20}$$

Corresponding Gradient of EU & RR

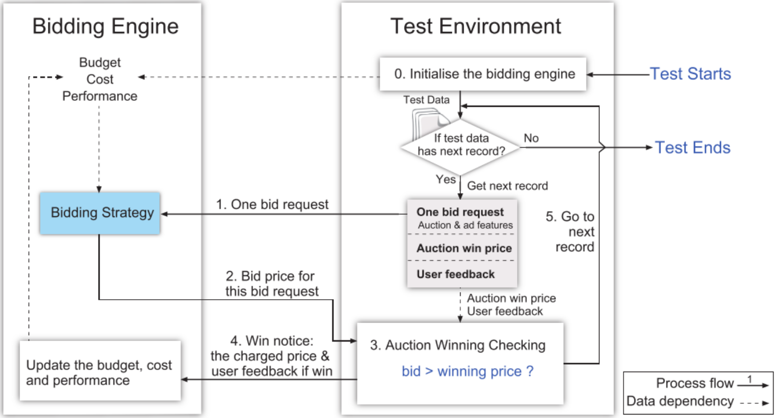
$$\frac{\partial R^{\text{EU}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = v^2 l (\sigma(\boldsymbol{\theta}^T \boldsymbol{x}) - y) \cdot \sigma(\boldsymbol{\theta}^T \boldsymbol{x})(1 - \sigma(\boldsymbol{\theta}^T \boldsymbol{x}))\boldsymbol{x} + \lambda \boldsymbol{\theta}, \tag{21}$$

$$\frac{\partial R^{\text{RR}}_{\boldsymbol{\theta}}(\boldsymbol{x}, y)}{\partial \boldsymbol{\theta}} = v l (\sigma(\boldsymbol{\theta}^T \boldsymbol{x}) - y)\boldsymbol{x} + \lambda \boldsymbol{\theta}. \tag{22}$$

Outline

1. User Response Prediction
   - Goal and Applications
   - State-of-the-art Methods
2. Case Study: Real-time Bidding
3. Optimize the Response Prediction
   - Problem Setup
   - Optimization for Campaign Performance
     - Expected Utility Model
     - Risk Return Model
   - Realization of Model
   - Links to Previous Work
4. Experiments
   - Evaluations

Evaluation Flow



Evaluation Measures

AUC

RMSE

Evaluation Measures

AUC

RMSE

profit = gain - cost

ROI = profit / cost

CTR = click # / impression #

eCPC = cost / click #

CPM = $10^3 \times$ cost / impression #

Dataset

iPinYou

[a] 64.75M bids, 19.5 imps, 14.79 clicks and 16K expense on 9 camps over 10 days.

[a] **http://data.computational-advertising.org**.

YOYI

[a] 443M imps, 362K clicks and 210K CNY expense over 8 days.

[a] **http://goo.gl/xaao4q**.

Compared Settings

User response prediction (truthful bidding function $b(\boldsymbol{x}) = v \cdot f(\boldsymbol{x})$)

CE - Cross entropy loss logistic regression

SE - Squared loss logistic regression

EU - Expected utility model

RR - Risk return model

Accuracy of CTR Estimation

Table: Regression performances over campaigns. AUC: the higher, the better. RMSE: the smaller, the better.

	AUC				RMSE ($\times 10^{-2}$)			
iPinYou	SE	CE	EU	RR	SE	CE	EU	RR
1458	.948	.987	.987	.977	3.01	1.94	2.42	2.32
2259	.542	.692	.674	.691	2.01	1.77	1.76	1.79
2261	.490	.569	.622	.619	1.84	1.68	1.71	1.68
2821	.511	.620	.608	.639	2.56	2.43	2.39	2.46
2997	.543	.610	.606	.608	5.98	5.82	5.84	5.82
3358	.863	.974	.970	.980	3.07	2.47	3.32	2.67
3386	.593	.768	.761	.778	2.95	2.84	3.32	2.85
3427	.634	.976	.976	.960	2.78	2.20	2.61	2.34
3476	.575	.957	.954	.950	2.50	2.32	2.39	2.33
Average	.633	.794	.795	**.800**	2.97	**2.61**	2.86	2.69
YOYI	.882	.891	**.912**	**.912**	11.9	11.7	11.8	**11.6**

Learning Curve

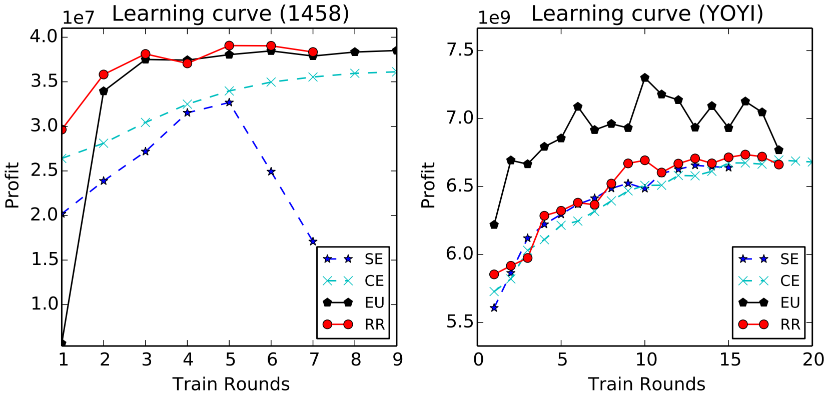


Figure: Training on iPinYou (left) and YOYI (right).

Campaign Profit Evaluation (baselines)

Table: Direct campaign profit over baselines.

	profit($\times 10^7$)		ROI	
iPinYou	SE	CE	SE	CE
1458	3.2	3.6	4.2	6.6
2259	-0.32	0.40	-0.080	0.18
2261	0.29	0.63	0.26	0.40
2821	0.11	0.08	0.21	0.023
2997	0.11	0.14	0.42	0.71
3358	1.76	2.4	5.4	5.2
3386	0.51	1.6	0.16	1.2
3427	0.33	2.9	0.11	3.4
3476	0.65	3.1	0.36	3.5
Average	0.74	**1.7**	1.2	**2.3**
YOYI	665.6	**669.5**	1.8	**1.9**

Campaign Profit Evaluation (baselines)

Table: Campaign profit improvement over baseline CE.

	Profit gain		ROI gain	
iPinYou	EU	RR	EU	RR
1458	7.10%	9.00%	233%	267%
2259	81.6%	99.3%	233%	472%
2261	26.3%	31.1%	44.4%	91.2%
2821	573%	615%	1334%	943%
2997	5.00%	0.700%	-3.60%	-11.4%
3358	1.70%	6.70%	77.1%	77.7%
3386	-1.20%	2.50%	20.6%	58.3%
3427	5.50%	8.70%	52.0%	175%
3476	4.20%	8.60%	16.0%	91.1%
YOYI	9.04%	0.600%	14.8%	2.11%
Average	+71.2%	+78.2%	+202%	+217%

Overall Statistics

	CTR ($\times 10^{-4}$)				eCPC			
iPinYou	SE	CE	EU	RR	SE	CE	EU	RR
1458	34	33	59	190	17	11	4.3	3.4
2259	3.3	3.6	3.7	5.8	303	235	172	136
2261	2.4	2.7	3.0	2.8	234	212	188	168
2821	5.5	5.9	4.8	7.0	116	137	105	112
2997	31	25	26	27	9.8	8.2	8.3	8.6
3358	51	41	69	61	18	19	12	12
3386	7.8	11	13	15	90	48	43	36
3427	7.2	25	29	72.8	98	25	17.3	10
3476	6.4	16	17	33.1	111	34	30	20
Average	16	18	25	**46**	110	81	64	**57**
YOYI	16	18	**26**	24	12.9	12.4	**11.3**	12
	CPM				Win Rate			
iPinYou	SE	CE	EU	RR	SE	CE	EU	RR
1458	57	37	25	65	0.22	0.24	0.13	.041
2259	100	84	64	78	0.89	0.63	0.44	0.24
2261	57	56	56	46	0.55	0.81	0.71	0.67
2821	63	80	50	78	0.12	0.63	0.48	0.45
2997	30	20	21	22	0.55	0.63	0.65	0.63
3358	92	77	80	70	0.11	0.20	0.11	0.13
3386	71	54	55	55	0.82	0.45	0.36	0.29
3427	70	60	49	75	0.75	0.26	0.22	.082
3476	71	55	50	65	0.49	0.31	0.31	0.15
Average	68	58	**50**	62	**0.50**	0.46	0.38	0.30
YOYI	**20**	23	29	30	**0.36**	0.30	0.22	0.22

Bidding Analysis

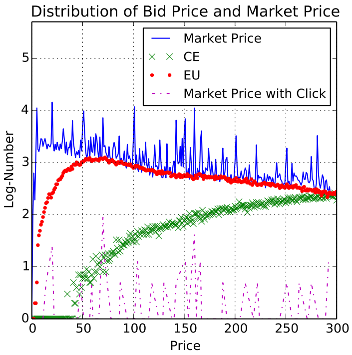


Figure: Analysis of bid price and market price distribution (iPinYou campaign 2259)

Online A/B Testing

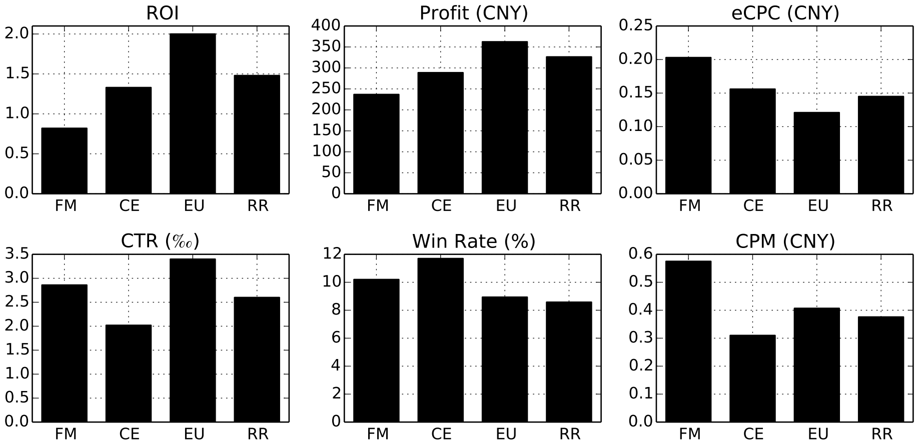


Figure: Online A/B testing results on YOYI PLUS.

Question

What if we optimize the total *Utility* w.r.t. the bidding function $b(\cdot)$?

What if considering the budget constraints?

Question

What if we optimize the total *Utility* w.r.t. the bidding function $b(\cdot)$?

What if considering the budget constraints?

Example

$$b^* = \arg\max_b \int_{\boldsymbol{x}} R_{\boldsymbol{\theta}}(\boldsymbol{x}, y; b, v, c, w) p_{\boldsymbol{x}}(\boldsymbol{x}) d\boldsymbol{x}, \tag{23}$$

$$s.t. \int_{\boldsymbol{x}} c(\boldsymbol{x}, y; b, w) p_{\boldsymbol{x}}(\boldsymbol{x}) d\boldsymbol{x} = B. \tag{24}$$

EM-fashion Optimization for $f_{\theta}(\boldsymbol{x})$ & $b(r)$

Input: Training set D, learning rate α, total budget B
Output: Optimal $b(\cdot)$ and $f_{\boldsymbol{\theta}}(\cdot)$

```
Initially set parameter θ and μ
while not converged do
    (E-Step)
    for each sample (x, y) ∈ D do
        Calculate the gradient via Eq. (13) or (14)
        Optimize θ with gradient descent
    end for
    (M-Step)
    Update bidding function b(·) via solving Eq. (28)
end while
```

Summary

Context (market price distribution and bid utility) contribute to response prediction.

We proposed:

- A novel market aware CTR estimation model;
- Discussion about the relation between our proposed model and the traditional learning models;
- A new dataset consisting over 87GB bidding logs for computational advertising research;
- A jointly optimization model for both $f(\boldsymbol{x})$ and $b(\cdot)$.

The newly proposed model can largely save budgets and earn more profits.

Derivation for Optimal Bidding Function

Once we fixed $f_{\boldsymbol{\theta}}(\boldsymbol{x})$, with the auction volume T and campaign budget B, we optimize ϕ in Eq. (12) as

$$\begin{aligned} \arg\max_{\phi} \; & T \int_r \int_0^{\phi v r} (vr - z) p_z(z) dz \cdot p_r(r) dr \\ \text{s.t. } \; & T \int_r \int_0^{\phi v r} z p_z(z) dz \cdot p_r(r) dr = B, \end{aligned} \tag{25}$$

where to simplify our notation, we substitute $f_{\boldsymbol{\theta}}(\boldsymbol{x})$ with its predicted CTR variable r. The Lagrangian $\mathcal{L}(\phi, \mu) =$

$$T \int_r \int_0^{\phi v r} \left[vr - (\mu + 1) z \right] p_z(z) dz \cdot p_r(r) dr + \mu B, \tag{26}$$

where μ is the Lagrangian multiplier. Taking the derivative equal to zero, we get that

$$\frac{\partial \mathcal{L}(\phi, \mu)}{\partial \phi} = 0 \;\Rightarrow\; \phi = \frac{1}{\mu + 1}. \tag{27}$$

Solve μ

To solve μ, we take the Lagrangian derivative w.r.t. to μ ant let it be zero, which obtains the constaint equation

$$T \int_{r} \int_{0}^{\frac{vr}{1+\mu}} z p_z(z) dz \; p_r(r) dr = B. \tag{28}$$

Jointly Optimization Results

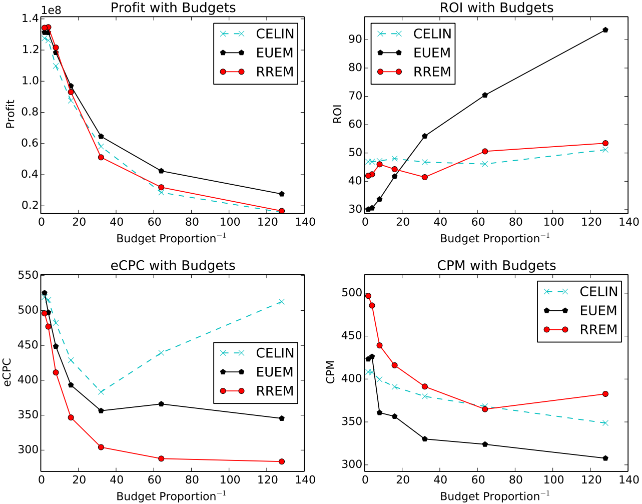


Figure: Performances with budgets on iPinYou.

Effectiveness of Jointly Optimization

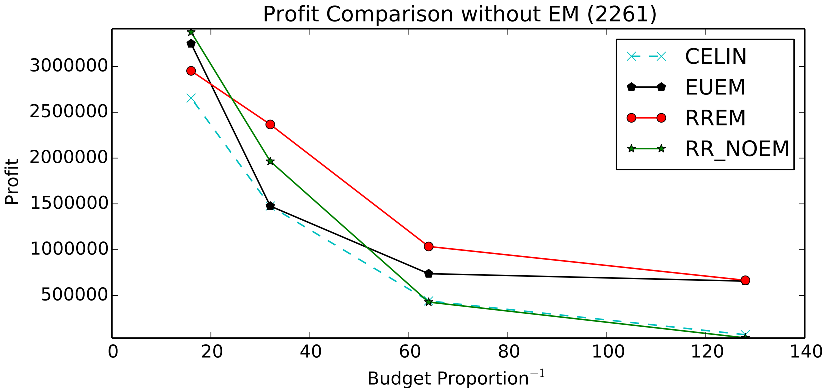


Figure: Joint optimization v.s. CTR optimization only.