User Response Learning for Directly Optimizing Campaign Performance in Display Advertising

Kan Ren\textsuperscript{1}  Weinan Zhang\textsuperscript{1}  Yifei Rong\textsuperscript{3}
Haifeng Zhang\textsuperscript{4}  Yong Yu\textsuperscript{1}  Jun Wang\textsuperscript{2}

\textsuperscript{1}Apex Data and Knowledge Management Lab
Shanghai Jiao Tong University

\textsuperscript{2}University College London & MediaGamma Ltd

\textsuperscript{3}YOYI Inc.

\textsuperscript{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

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
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

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
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
Objective Function

Squared Error

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

Cross Entropy

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

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
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

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
## Notations and descriptions

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>The true label of user response.</td>
</tr>
<tr>
<td>$x$</td>
<td>The bid request represented by its features.</td>
</tr>
<tr>
<td>$\theta$</td>
<td>The parameter of CTR estimation function.</td>
</tr>
<tr>
<td>$f_\theta(x)$</td>
<td>the CTR estimation function to learn.</td>
</tr>
<tr>
<td>$b(f_\theta(x))$</td>
<td>The bid price determined by the estimated CTR, b for short.</td>
</tr>
<tr>
<td>$R_\theta(\cdot)$</td>
<td>The utility function.</td>
</tr>
</tbody>
</table>
Market Modeling

Market Price

The second highest price proposed during an RTB auction (2\textsuperscript{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.)

$$w(b) = \int_0^b p_z(z)dz.$$  (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.)
\[ w(b) = \int_0^b p_z(z)dz. \] (1)

Expected Cost under $2^{nd}$ Price Auction
\[ c(b) = \int_0^b zp_z(z)dz, \int_0^b zp_z(z)dz \] (2)
Objective Function

\[
\theta^* = \arg \max_{\theta} \int_x R_{\theta}(x, y; b, v, c, w)p_x(x)dx.
\]

(3)

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

Constant click value \(v\) limits the max bid.

We will propose two variants of \(R_{\theta}(\cdot)\).
Expected Utility Model

Expected Utility (EU)

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

(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

\[
\sum_{(x,y) \in D} R_{\theta}^{EU}(x, y) = \sum_{(x,y) \in D} [vy - c(b(f_{\theta}(x)))] \cdot w(b(f_{\theta}(x)))
\]

\[
= \sum_{(x,y) \in D} \left[ vy - \frac{\int_0^{b(f_{\theta}(x))} z \cdot p_z(z)dz}{\int_0^{b(f_{\theta}(x))} p_z(z)dz} \right] \cdot \int_0^{b(f_{\theta}(x))} p_z(z)dz
\]

\[
= \sum_{(x,y) \in D} \int_0^{b(f_{\theta}(x))} (vy - z) \cdot p_z(z)dz. \quad (5)
\]

\(^2\)O. Chapelle. WWW 2015.
Optimal Parameter

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

\[
\theta^{EU} = \arg \min_{\theta} - \sum_{(x,y) \in D} R^{EU}_{\theta}(x, y) + \frac{\lambda}{2} \| \theta \|_2^2
\]

\[
= \arg \min_{\theta} \sum_{x} \int_{0}^{b(f_{\theta}(x))} (z - vy) \cdot p_{z}(z) dz + \frac{\lambda}{2} \theta^{T}\theta.
\]

where the optimal value of \( \theta \) is obtained by taking a gradient descent algorithm.
Gradient of EU

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

$$\frac{\partial R_{\theta}^{EU}(x, y)}{\partial \theta} = \frac{\text{bid error}}{\text{market sensitivity}} \cdot \frac{\partial b(f_{\theta}(x))}{\partial f_{\theta}(x)} \cdot \frac{\partial f_{\theta}(x)}{\partial \theta} + \lambda \theta.$$  

(7)

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

Risk Return (RR)

\[ R_{\theta}^{RR}(x, y) = \left( \frac{vy}{z} - \frac{v(1 - y)}{v - z} \right) \cdot w(b(f_\theta(x))). \]  

(8)
Objective of RR

\[
\theta^{RR} = \arg \min_{\theta} - \sum_{(x, y) \in D} \mathcal{R}_{\theta}^{RR}(x, y) + \frac{\lambda}{2} \| \theta \|^2
\]

\[
= \arg \min_{\theta} - \sum_{(x, y) \in D} \int_0^{b(f_\theta(x))} \left( \frac{vy}{z} - \frac{v(1 - y)}{v - z} \right) \cdot p_z(z) dz \]

\[+ \frac{\lambda}{2} \theta^T \theta. \tag{9}\]
Gradient of RR

\[
\frac{\partial R_{\theta}^{RR}(x, y)}{\partial \theta} = \left( -\frac{vy}{b(f_{\theta}(x))} + \frac{v(1 - y)}{v - b(f_{\theta}(x))} \right) \cdot p_z(b(f_{\theta}(x))) \\
\cdot \frac{\partial b(f_{\theta}(x))}{\partial f_{\theta}(x)} \frac{\partial f_{\theta}(x)}{\partial \theta} + \lambda \theta.
\]
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
Question

Gradient of EU

\[
\frac{\partial R_{\theta}^{EU}(x, y)}{\partial \theta} = \left( b(f_\theta(x)) - vy \right) \cdot \left( \frac{\partial b(f_\theta(x))}{\partial f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta} \right) + \lambda \theta.
\]

Gradient of RR

\[
\frac{\partial R_{\theta}^{RR}(x, y)}{\partial \theta} = \left( \frac{-vy}{b(f_\theta(x))} + \frac{v(1 - y)}{v - b(f_\theta(x))} \right) \cdot \left( \frac{\partial b(f_\theta(x))}{\partial f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta} \right) + \lambda \theta.
\]
Question

Gradient of EU

\[
\frac{\partial R_{\theta}^{EU}(x, y)}{\partial \theta} = \left( b(f_\theta(x)) - vy \right) \cdot p_z(b(f_\theta(x))) \cdot \frac{\partial b(f_\theta(x))}{\partial f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta} + \lambda \theta.
\]

Gradient of RR

\[
\frac{\partial R_{\theta}^{RR}(x, y)}{\partial \theta} = \left( -\frac{vy}{b(f_\theta(x))} + \frac{v(1 - y)}{v - b(f_\theta(x))} \right) \cdot p_z(b(f_\theta(x))) \cdot \frac{\partial b(f_\theta(x))}{\partial f_\theta(x)} \cdot \frac{\partial f_\theta(x)}{\partial \theta} + \lambda \theta.
\]
Realization of Model

Response Prediction Model $f_\theta(x)$

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

Linear Bidding Strategy

$$b(f_\theta(x)) \equiv \phi \cdot v \cdot f_\theta(x),$$ \hspace{1cm} (12)

where $\phi$ is the scaling parameter.
Linear Gradient

Linear Gradient of EU

\[
\frac{\partial R^{\text{EU}}_{\theta}(x, y)}{\partial \theta} = \phi v^2 (\phi \sigma(\theta^T x) - y) \cdot p_z(b(f_\theta(x))) \cdot \\
\sigma(\theta^T x)(1 - \sigma(\theta^T x))x + \lambda \theta. \tag{13}
\]

Linear Gradient of RR

\[
\frac{\partial R^{\text{RR}}_{\theta}(x, y)}{\partial \theta} = \phi v \left( - \frac{y}{\phi \sigma(\theta^T x)} + \frac{1 - y}{1 - \phi \sigma(\theta^T x)} \right) \cdot \\
p_z(b(f_\theta(x))) \cdot \sigma(\theta^T x)(1 - \sigma(\theta^T x))x + \lambda \theta. \tag{14}
\]
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
Traditional Logistic Regression

Squared Error LR (SE)

\[ \mathcal{L}_{\theta}^{\text{SE}}(x, y) = \frac{1}{2} (y - \sigma(\theta^T x))^2, \]

\[ \frac{\partial \mathcal{L}_{\theta}^{\text{SE}}(x, y)}{\partial \theta} = (\sigma(\theta^T x) - y)\sigma(\theta^T x)(1 - \sigma(\theta^T x))x. \] (15)

Cross Entropy LR (CE)

\[ \mathcal{L}_{\theta}^{\text{CE}}(x, y) = -y \log \sigma(\theta^T x) - (1 - y) \log(1 - \sigma(\theta^T x)), \]

\[ \frac{\partial \mathcal{L}_{\theta}^{\text{CE}}(x, y)}{\partial \theta} = (\sigma(\theta^T x) - y)x. \] (16)
Discussion: Truthful Bidding Simplification

Simplification: Truthful Bidding, $\phi = 1$

$$b(f_\theta(x)) = v \cdot f_\theta(x).$$ (17)

Corresponding Gradient of EU & RR

$$\frac{\partial R^{EU}_\theta(x, y)}{\partial \theta} = v^2 (\sigma(\theta^T x) - y) \cdot p_z(b(f_\theta(x))) \sigma(\theta^T x)(1 - \sigma(\theta^T x))x + \lambda \theta,$$

$$\frac{\partial R^{RR}_\theta(x, y)}{\partial \theta} = v (\sigma(\theta^T x) - y) p_z(b(f_\theta(x)))x + \lambda \theta,$$ (19)

EU & RR have one more component (market sensitivity) than SE & CE, respectively!
Discussion: Uniform Market Price Distribution Simplification

Simplification: Uniform Market Price Distribution

\[ p_z(z) = I. \quad (20) \]

Corresponding Gradient of EU & RR

\[ \frac{\partial R_{\theta}^{EU}(x, y)}{\partial \theta} = v^2 I(\sigma(\theta^T x) - y) \cdot \sigma(\theta^T x)(1 - \sigma(\theta^T x))x + \lambda \theta, \quad (21) \]

\[ \frac{\partial R_{\theta}^{RR}(x, y)}{\partial \theta} = vl(\sigma(\theta^T x) - y)x + \lambda \theta. \quad (22) \]
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

**Bidding Engine**
- Budget
- Cost
- Performance
- Bidding Strategy
- Update the budget, cost and performance

**Test Environment**
1. One bid request
2. Bid price for this bid request
4. Win notice: the charged price & user feedback if win

**One bid request**
- Auction & ad features
- Auction win price
- User feedback

3. Auction Winning Checking
   - bid > winning price?

5. Go to next record

Process flow

Test Starts

Test Ends

Data dependency - - - -
Evaluation Measures

AUC
RMSE
Evaluation Measures

AUC
RMSE

profit = gain - cost
ROI = profit / cost
CTR = click # / impression #
eCPC = cost / click #
CPM = $10^3 \times \frac{\text{cost}}{\text{impression} \#}$
Dataset

**iPinYou**

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


**YOYI**

^443M imps, 362K clicks and 210K CNY expense over 8 days.

Compared Settings

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

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.

<table>
<thead>
<tr>
<th>iPinYou</th>
<th>AUC</th>
<th>RMSE ($\times 10^{-2}$)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>CE</td>
<td>EU</td>
</tr>
<tr>
<td>1458</td>
<td>0.948</td>
<td>0.987</td>
<td>0.987</td>
</tr>
<tr>
<td>2259</td>
<td>0.542</td>
<td>0.692</td>
<td>0.674</td>
</tr>
<tr>
<td>2261</td>
<td>0.490</td>
<td>0.569</td>
<td>0.622</td>
</tr>
<tr>
<td>2821</td>
<td>0.511</td>
<td>0.620</td>
<td>0.608</td>
</tr>
<tr>
<td>2997</td>
<td>0.543</td>
<td>0.610</td>
<td>0.606</td>
</tr>
<tr>
<td>3358</td>
<td>0.863</td>
<td>0.974</td>
<td>0.970</td>
</tr>
<tr>
<td>3386</td>
<td>0.593</td>
<td>0.768</td>
<td>0.761</td>
</tr>
<tr>
<td>3427</td>
<td>0.634</td>
<td>0.976</td>
<td>0.976</td>
</tr>
<tr>
<td>3476</td>
<td>0.575</td>
<td>0.957</td>
<td>0.954</td>
</tr>
<tr>
<td>Average</td>
<td>0.633</td>
<td>0.794</td>
<td>0.795</td>
</tr>
<tr>
<td>YOYI</td>
<td>0.882</td>
<td>0.891</td>
<td><strong>0.912</strong></td>
</tr>
</tbody>
</table>
Learning Curve

Figure: Training on iPinYou (left) and YOYI (right).
## Campaign Profit Evaluation (baselines)

**Table:** Direct campaign profit over baselines.

<table>
<thead>
<tr>
<th></th>
<th>profit($\times 10^7$)</th>
<th></th>
<th>ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>CE</td>
<td>SE</td>
</tr>
<tr>
<td>iPinYou</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1458</td>
<td>3.2</td>
<td>3.6</td>
<td>4.2</td>
</tr>
<tr>
<td>2259</td>
<td>-0.32</td>
<td>0.40</td>
<td>-0.08</td>
</tr>
<tr>
<td>2261</td>
<td>0.29</td>
<td>0.63</td>
<td>0.26</td>
</tr>
<tr>
<td>2821</td>
<td>0.11</td>
<td>0.08</td>
<td>0.21</td>
</tr>
<tr>
<td>2997</td>
<td>0.11</td>
<td>0.14</td>
<td>0.42</td>
</tr>
<tr>
<td>3358</td>
<td>1.76</td>
<td>2.4</td>
<td>5.4</td>
</tr>
<tr>
<td>3386</td>
<td>0.51</td>
<td>1.6</td>
<td>0.16</td>
</tr>
<tr>
<td>3427</td>
<td>0.33</td>
<td>2.9</td>
<td>0.11</td>
</tr>
<tr>
<td>3476</td>
<td>0.65</td>
<td>3.1</td>
<td>0.36</td>
</tr>
<tr>
<td>Average</td>
<td>0.74</td>
<td>1.7</td>
<td>1.2</td>
</tr>
<tr>
<td>YOYI</td>
<td>665.6</td>
<td>669.5</td>
<td>1.8</td>
</tr>
</tbody>
</table>
## Campaign Profit Evaluation (baselines)

### Table: Campaign profit improvement over baseline CE.

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

<table>
<thead>
<tr>
<th></th>
<th>SE</th>
<th>CE</th>
<th>EU</th>
<th>RR</th>
<th>SE</th>
<th>CE</th>
<th>EU</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTR ($\times 10^{-4}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
<td>11</td>
<td>4.3</td>
<td>3.4</td>
</tr>
<tr>
<td>1458</td>
<td>34</td>
<td>33</td>
<td>59</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2259</td>
<td>3.3</td>
<td>3.6</td>
<td>3.7</td>
<td>5.8</td>
<td>303</td>
<td>235</td>
<td>172</td>
<td>136</td>
</tr>
<tr>
<td>2261</td>
<td>2.4</td>
<td>2.7</td>
<td>3.0</td>
<td>2.8</td>
<td>234</td>
<td>212</td>
<td>188</td>
<td>168</td>
</tr>
<tr>
<td>2821</td>
<td>5.5</td>
<td>5.9</td>
<td>4.8</td>
<td>7.0</td>
<td>116</td>
<td>137</td>
<td>105</td>
<td>112</td>
</tr>
<tr>
<td>2997</td>
<td>31</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>9.8</td>
<td>8.2</td>
<td>8.3</td>
<td>8.6</td>
</tr>
<tr>
<td>3358</td>
<td>51</td>
<td>41</td>
<td>69</td>
<td>61</td>
<td>18</td>
<td>19</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>3386</td>
<td>7.8</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>90</td>
<td>48</td>
<td>43</td>
<td>36</td>
</tr>
<tr>
<td>3427</td>
<td>7.2</td>
<td>25</td>
<td>29</td>
<td>72.8</td>
<td>98</td>
<td>25</td>
<td>17.3</td>
<td>10</td>
</tr>
<tr>
<td>3476</td>
<td>6.4</td>
<td>16</td>
<td>17</td>
<td>33.1</td>
<td>111</td>
<td>34</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Average</td>
<td>16</td>
<td>18</td>
<td>25</td>
<td>46</td>
<td>110</td>
<td>81</td>
<td>64</td>
<td>57</td>
</tr>
<tr>
<td>YOYI</td>
<td>16</td>
<td>18</td>
<td>26</td>
<td>24</td>
<td>12.9</td>
<td>12.4</td>
<td>11.3</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SE</th>
<th>CE</th>
<th>EU</th>
<th>RR</th>
<th>SE</th>
<th>CE</th>
<th>EU</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1458</td>
<td>57</td>
<td>37</td>
<td>25</td>
<td>65</td>
<td>0.22</td>
<td>0.24</td>
<td>0.13</td>
<td>0.041</td>
</tr>
<tr>
<td>2259</td>
<td>100</td>
<td>84</td>
<td>64</td>
<td>78</td>
<td>0.89</td>
<td>0.63</td>
<td>0.44</td>
<td>0.24</td>
</tr>
<tr>
<td>2261</td>
<td>57</td>
<td>56</td>
<td>56</td>
<td>46</td>
<td>0.55</td>
<td>0.81</td>
<td>0.71</td>
<td>0.67</td>
</tr>
<tr>
<td>2821</td>
<td>63</td>
<td>80</td>
<td>50</td>
<td>78</td>
<td>0.12</td>
<td>0.63</td>
<td>0.48</td>
<td>0.45</td>
</tr>
<tr>
<td>2997</td>
<td>30</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>0.55</td>
<td>0.63</td>
<td>0.65</td>
<td>0.63</td>
</tr>
<tr>
<td>3358</td>
<td>92</td>
<td>77</td>
<td>80</td>
<td>70</td>
<td>0.11</td>
<td>0.20</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>3386</td>
<td>71</td>
<td>54</td>
<td>55</td>
<td>55</td>
<td>0.82</td>
<td>0.45</td>
<td>0.36</td>
<td>0.29</td>
</tr>
<tr>
<td>3427</td>
<td>70</td>
<td>60</td>
<td>49</td>
<td>75</td>
<td>0.75</td>
<td>0.26</td>
<td>0.22</td>
<td>0.082</td>
</tr>
<tr>
<td>3476</td>
<td>71</td>
<td>55</td>
<td>50</td>
<td>65</td>
<td>0.49</td>
<td>0.31</td>
<td>0.31</td>
<td>0.15</td>
</tr>
<tr>
<td>Average</td>
<td>68</td>
<td>58</td>
<td>50</td>
<td>62</td>
<td>0.50</td>
<td>0.46</td>
<td>0.38</td>
<td>0.30</td>
</tr>
<tr>
<td>YOYI</td>
<td>20</td>
<td>23</td>
<td>29</td>
<td>30</td>
<td>0.36</td>
<td>0.30</td>
<td>0.22</td>
<td>0.22</td>
</tr>
</tbody>
</table>
**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

$$b^* = \arg \max_b \int_x R_\theta(x, y; b, v, c, w)p_x(x)dx,$$

(23)

$$\text{s.t. } \int_x c(x, y; b, w)p_x(x)dx = B.$$ 

(24)
EM-fashion Optimization for $f_\theta(x)$ & $b(r)$

**Input:** Training set $D$, learning rate $\alpha$, total budget $B$

**Output:** Optimal $b(\cdot)$ and $f_\theta(\cdot)$

1. Initially set parameter $\theta$ and $\mu$
2. **while** not converged **do**
   3. (E-Step)
   4. **for** each sample $(x, y) \in D$ **do**
   5. Calculate the gradient via Eq. (13) or (14)
   6. Optimize $\theta$ with gradient descent
   7. **end for**
   8. (M-Step)
   9. Update bidding function $b(\cdot)$ via solving Eq. (28)
10. **end while**
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 $f(x)$ and $b(\cdot)$.

The newly proposed model can largely save budgets and earn more profits.
Derivation for Optimal Bidding Function

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

$$\arg \max_\phi T \int_r \int_0^{\phi vr} (vr - z)p_z(z)dz \cdot p_r(r)dr$$

s.t. $T \int_r \int_0^{\phi vr} zp_z(z)dz \cdot p_r(r)dr = B,$

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

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

where $\mu$ is the Lagrangian multiplier. Taking the derivative equal to zero, we get that

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

To solve $\mu$, we take the Lagrangian derivative w.r.t. to $\mu$ and let it be zero, which obtains the constraint equation

$$T \int_r \int_0^{\frac{vr}{1+\mu}} zp_z(z)dz \ p_r(r)dr = B.$$  \hspace{1cm} (28)
Jointly Optimization Results

Figure: Performances with budgets on iPinYou.
Effectiveness of Jointly Optimization

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