Unbiased Lift–based Bidding System

Daisuke Moriwaki, Yuta Hayakawa, Isshu Munemasa
CyberAgent, Inc.
{moriwaki_daisuke, hayakawa_yuta, munemasa_isshu}@cyberagent.co.jp

Yuta Saito
Tokyo Institute of Technology
saito.y.bj@m.titech.ac.jp

Akira Matsui
University of Southern California
amatui@usc.edu
Background

- DSPs (Demand-Side Platforms) Buy adslot through Real-Time Bidding

Unbiased Lift-based Bidding System, AdKDD2020
Background

- DSPs (Demand-Side Platforms) Buy adslot through Real-Time Bidding
- The prevalent bidding strategy is *Performance-based* or *Value-based* bidding

\[
\text{bid} = \text{CPC} \cdot eCTR \cdot \alpha
\]

or

\[
\text{bid} = \text{CPA} \cdot eCTR \cdot eCVR \cdot \alpha
\]
Background

- DSPs (Demand-Side Platforms) Buy adslot through Real-Time Bidding
- The prevalent bidding strategy is *Performance-based* or *Value-based* bidding
  
  \[ \text{bid} = \text{CPC} \cdot \text{eCTR} \cdot \alpha \]
  
  or
  
  \[ \text{bid} = \text{CPA} \cdot \text{eCTR} \cdot \text{eCVR} \cdot \alpha \]

- Performance-based bidding is optimal under current billing system CPC, CPA because it maximizes DSP’s profit
Background

- While the strategy is consistent with the DSP’s objective, it is NOT always the optimal strategy for the clients (advertisers)
Background

- While the strategy is consistent with the DSP’s objective, it is NOT always the optimal strategy for the clients (advertisers)

- Example
  - A user may purchase a good without ad while the advertiser pay DSP fee for the ad
  - A user may stop buying a good because of an ad while the DSP is not given any penalty
Background

- More specifically, suppose DSP bids adslot for two potential customers

\[
\begin{align*}
\mathbb{E}[Y|ad] &= 0.8 \\
\mathbb{E}[Y|\text{no } ad] &= 0.7 \\
\mathbb{E}[Y|ad] &= 0.2 \\
\mathbb{E}[Y|\text{no } ad] &= 0.0
\end{align*}
\]

<table>
<thead>
<tr>
<th>Performance-based bidder</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal bidder</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

Unbiased Lift–based Bidding System, AdKDD2020
Background

- More specifically, suppose DSP bids adslot for two potential customers
- Performance-based bidder bids high for user with high conversion rate given ad

<table>
<thead>
<tr>
<th>Performance-based bidder</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal bidder</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>
Background

- More specifically, suppose DSP bids adslot for two potential customers
- Performance–based bidder bids high for user with high conversion rate given ad
- Optimal bidder should bids high for user with high lift

\[
\begin{align*}
\mathbb{E}[Y|\text{ad}] &= 0.8 \\
\mathbb{E}[Y|\text{no ad}] &= 0.7
\end{align*}
\]

\[
\begin{align*}
\mathbb{E}[Y|\text{ad}] &= 0.2 \\
\mathbb{E}[Y|\text{no ad}] &= 0.0
\end{align*}
\]

<table>
<thead>
<tr>
<th>Performance-based bidder</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal bidder</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>
Proposed Method

- *Lift-based bidding* is a bidding strategy that the bid price is based on the predicted lift $\tau$

\[
bid_{i,s(a)} = \nu \tau(s(a) | x_i) \alpha
\]
Proposed Method

- *Lift-based bidding* is a bidding strategy that the bid price is based on the predicted lift \( \tau \)

\[
bid_{i,s(a)} = v \tau(s(a) | x_i) \alpha
\]

- \( v \) is reward for a unit of conversion lift
Proposed Method

- *Lift-based bidding* is a bidding strategy that the bid price is based on the predicted lift $\tau$

\[
bid_{i,s(a)} = v\tau(s(a) \mid x_i)\alpha
\]

- $v$ is reward for a unit of conversion lift
- $\alpha$ is a parameter to adjust to RTB environment (competitiveness, # bid requests etc.)
Proposed Method

- *Lift–based bidding* is a bidding strategy that the bid price is based on the predicted lift $\tau$

\[
bid_{i,s(a)} = v\tau(s(a) \mid \mathbf{x}_i)\alpha
\]

- $v$ is reward for a unit of conversion lift
- $\alpha$ is a parameter to adjust to RTB environment (competitiveness, # bid requests etc.)
- $\tau$ is calculated as the difference between two potential outcomes
  - $s(a)$: # ad exposure for ad $a$
    \[
    \tau(s(a) \mid \mathbf{x}_i) = \mathbb{E}[y_i \mid s(a), \mathbf{x}_i] - \mathbb{E}[y_i \mid s(a) - 1, \mathbf{x}_i]
    \]
    - Expected outcome with additional ad
    - Expected outcome without additional ad
Proposed Method

- **Lift–based bidding** is a bidding strategy that the bid price is based on the predicted lift $\tau$

  $$\text{bid}_{i,s(a)} = v\tau(s(a) \mid x_i)\alpha$$

- $v$ is reward for a unit of conversion lift
- $\alpha$ is a parameter to adjust to RTB environment (competitiveness, # bid requests etc.)
- $\tau$ is calculated as the difference between two *potential outcomes*
  - $s(a)$: # ad exposure for ad $a$
    $$\tau(s(a) \mid x_i) = \mathbb{E}[y_i \mid s(a), x_i] - \mathbb{E}[y_i \mid s(a) - 1, x_i]$$

- The basic idea is similar to Xu et al. (AAAI’16)
Unbiased Lift–Effect Prediction

- Each potential outcome is predicted by the corresponding predictor $f$

$$
\hat{\tau}(s(a) \mid x_i) = f^{s(a)}(x_i) - f^{s(a)-1}(x_i)
$$
Unbiased Lift–Effect Prediction

- Each potential outcome is predicted by the corresponding predictor $f$:
  $$\hat{\tau}(s(a) \mid x_i) = f^{s(a)}(x_i) - f^{s(a)-1}(x_i)$$

- Challenges
  - Ideal estimation needs data from population distribution $p(x, y, s(a))$
  - In reality, one of the potential outcomes is available
  - Assignment of ads is biased due to past bidding strategy (e.g. user with higher bid price are more likely to see ad)
  - Naive ERM (Empirical Risk Minimization) leads to a biased estimator
Unbiased Lift–Effect Prediction

- Each potential outcome is predicted by the corresponding predictor $f$

\[ \hat{\tau}(s(a) \mid x_i) = f^{s(a)}(x_i) - f^{s(a)-1}(x_i) \]

- Challenges
  - Ideal estimation needs data from population distribution $p(x, y, s(a))$
  - In reality, one of the potential outcomes is available.
  - Assignment of ads is biased due to past bidding strategy (e.g. user with higher bid price are more likely to see ad)
  - Naive ERM (Empirical Risk Minimization) leads to biased estimator

- Solution
  - IPS (Inverse Propensity Score) estimation to debias the estimation
Unbiased Lift–Effect Prediction

More specifically,

- Ideal loss function is defined as expectation of loss over population distribution

\[ L_{\text{ideal}}(f^{s(a)}) = \mathbb{E}_{p(x,y)}[l(y(s(a)), f^{s(a)}(x))] \]
Unbiased Lift–Effect Prediction

More specifically,

- Ideal loss function is defined as expectation of loss over population distribution

\[ \mathcal{L}_{\text{ideal}}(f^{s(a)}) = \mathbb{E}_{p(x,y)}[l(y(s(a)), f^{s(a)}(x))] \]

- A naive ERM estimation minimize loss over realized data which is biased since distribution of observed data is not the population distribution \( p(x,y|s(a)) \neq p(x,y) \)

\[ f^{s(a)}_{\text{ERM}} = \arg \min_{f^{s(a)}} \hat{\mathcal{L}}_{\text{ERM}}(f^{s(a)}) = \arg \min_{f^{s(a)}} \frac{1}{n_{s(a)}} \sum_{i \in D_{s(a)}} l(y_i^{\text{obs}}, f^{s(a)}(x_i)) \]
Unbiased Lift–Effect Prediction

More specifically,

- Ideal loss function is defined as expectation of loss over population distribution

\[ \mathcal{L}_{\text{ideal}}(f^{s(a)}) = \mathbb{E}_{p(x,y)}[l\left(y(s(a)), f^{s(a)}(x)\right)] \]

- A naive ERM estimation minimize loss over realized data which is biased since distribution of observed data is not the population distribution \( p(x,y|s(a)) \neq p(x,y) \)

\[ f_{\text{ERM}}^{s(a)} = \arg \min_{f^{s(a)}} \hat{\mathcal{L}}_{\text{ERM}}(f^{s(a)}) = \arg \min_{f^{s(a)}} \frac{1}{n_{s(a)}} \sum_{i \in D_{s(a)}} l(y_{i}^{obs}, f^{s(a)}(x_{i})) \]

- IPS \( e_{s(a)}(x_{i}) \) corrects for the bias of observed data by reweighting loss

\[ \hat{\mathcal{L}}_{\text{IPS}}(f^{s(a)}) = \frac{1}{n} \sum_{i \in D_{s(a)}} \frac{1}{e_{s(a)}(x_{i})} l(y_{i}^{obs}, f^{s(a)}(x_{i})) \]
Implementation

- A unique DSP/DMP for O2O marketing provided by CyberAgent
Implementation

- A unique DSP/DMP for O2O marketing provided by CyberAgent
  - Target potential customers with behavior and deliver ads
Implementation

- A unique DSP/DMP for O2O marketing provided by CyberAgent
  - Target potential customers with behavior and deliver ads
  - Tracking offline store visits using location data (e.g. GPS)
Implementation

- A unique DSP/DMP for O2O marketing provided by CyberAgent
  - Target potential customers with behavior and deliver ads
  - Tracking offline store visits using location data (e.g. GPS)
  - Measuring causal effect by A/B testing
Unbiased Lift-based bidding system

- Bid request with device id and feature \( x \)
- Check past impression record
- Set pre-calculated \( \tau \)
- Calculate bid price = \( v \cdot \tau \cdot \alpha \)
- Updated \( \alpha \) to pace budget

\[
\tau(s(a) | x_i)
\]

Real-time update
Historical data
Bid request
Stream data
Impression count
Cost rate \( \alpha \)
Budget spending
PID control
Real-time Bidding
Update
Unbiased Lift-based bidding system

• Bid request with device id and feature x
• Check past impression record
• Set pre-calculated $\tau$
• Calculate bid price = $v \cdot \tau \cdot \alpha$
• Updated $\alpha$ to pace budget
Unbiased Lift-based bidding system

- Bid request with device id and feature $x$
- **Check past impression record**
- Set pre-calculated $\tau$
- Calculate bid price $= v \cdot \tau \cdot \alpha$
- Updated $\alpha$ to pace budget
Unbiased Lift-based bidding system

- Bid request with device id and feature $x$
- Check past impression record
- Set pre-calculated $\tau$
- Calculate bid price $= v \cdot \tau \cdot \alpha$
- Updated $\alpha$ to pace budget
Unbiased Lift-based bidding system

- Bid request with device id and feature x
- Check past impression record
- Set pre-calculated $\tau$
- Calculate bid price $= v \cdot \tau \cdot \alpha$
- Updated $\alpha$ to pace budget
Unbiased Lift-based bidding system

- Bid request with device id and feature $x$
- Check past impression record
- Set pre-calculated $\tau$
- Calculate bid price $= v \cdot \tau \cdot \alpha$
- Updated $\alpha$ to pace budget

\[
\tau(s(a) | x_i) = \text{Cost rate } \alpha
\]

Unbiased Lift-based Bidding System, AdKDD2020
Model Training

- Data extraction
  - Ad exposure
  - Feature
  - Response
Model Training

- Data extraction
  - Ad exposure
  - Feature
  - Response
- User-level data
  - Use all the data
  - Consistent with debias
Model Training

- Data extraction
  - Ad exposure
  - Feature
  - Response

- User-level data
  - Use all the data
  - Consistent with debias

- Models
  - First stage: IPS estimator $e(x)$
  - Second stage: outcome predictors $f_s$
  - Both models use XGBoost

<table>
<thead>
<tr>
<th>user</th>
<th>impression</th>
<th>conversion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Online Experiment

- Evaluate models in real advertising campaign
  - An electronics retailer (like Best Buy) in Japan promotes app downloads and visits to real stores with display ads
Online Experiment

- Evaluate models in real advertising campaign
  - An electronics retailer (like Best Buy) in Japan promotes app downloads and visits to real stores with display ads
- Lift-based bidder vs. Existing bidder
  - Existing bidder maximizes performance
Online Experiment

- Evaluate models in real advertising campaign
  - An electronics retailer (like Best Buy) in Japan promotes app downloads and visits to real stores with display ads
- Lift-based bidder vs. Existing bidder
  - Existing bidder maximizes performance
- Split users and budget into two groups
  - Assign 88% to existing algorithm for business reason
Results: ad performance

- Impressions and reach rate (# users with impression/#users) increased
- Clicks decreased
- No significant difference for conversion rate (CVR)

<table>
<thead>
<tr>
<th></th>
<th>Lift-based</th>
<th>Performance-based</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>impression-per-user</td>
<td>1.28</td>
<td>1.0</td>
<td>0.28***</td>
</tr>
<tr>
<td>clicks-per-user</td>
<td>0.54</td>
<td>1.0</td>
<td>-0.46**</td>
</tr>
<tr>
<td>reach rate</td>
<td>1.71</td>
<td>1.0</td>
<td>0.71***</td>
</tr>
</tbody>
</table>
Results: Business KPIs

- Besides the same CVR, cost has been remarkably decreased
- Lift-based bidding archived the same results with less than 30% of the money that existing bidder spent!

<table>
<thead>
<tr>
<th></th>
<th>Lift-based</th>
<th>Performance-based</th>
<th>business impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost-per-impression</td>
<td>0.27</td>
<td>1.0</td>
<td>73% saved!</td>
</tr>
<tr>
<td>Cost-per-reach</td>
<td>0.24</td>
<td>1.0</td>
<td>76% saved!</td>
</tr>
<tr>
<td>Cost-per-visit (CPA)</td>
<td>0.3</td>
<td>1.0</td>
<td>70% saved!</td>
</tr>
</tbody>
</table>
Discussion

- Lift-based algorithm
  - bid much lower bid price.
  - very stable
Discussion

- Lift-based algorithm
  - bid much lower bid price.
  - very stable

- Possible reasons
  - Other DSPs also follow performance-based algorithm. very competitive auctions
  - In contrast, lift-based bidder participates less competitive auctions. Win easily
Conclusion and Future Study

- Propose unbiased lift-based bidding system
- Debias lift-effect prediction with IPS
- Deployed in real environment
- Online experiment
- Future work includes
  - More real-world experiments
  - Post-hoc analysis
Thank you!