Intelligence on Kakao Advertising
beyond state-of-the-art

ben.hur@kakao corp.com
Nothing is certain but death and taxes.
- Benjamin Franklin
Nothing is certain but death and taxes.

“AD”.

- Benjamin Franklin
** Visit-to-Impression is done within 1~200ms for at least ten-thousand requests per second.

SSP: Supply-side platform  
DSP: Demand-side platform  
DMP: Data management platform  
MAT: Mobile app tracking  
AdX: Ad exchange  
RTB: Real-time bidding  

AdTech Eco (AdX/RTB/Programmatic Buying)
Different dreams among \{\textbf{Audience}, \textbf{Advertiser}, \textbf{Publisher}, \textbf{Platform}\}
Balance between **Revenue** and **Relevance**

- **Audience**: Provide relevant and interesting information
- **Advertiser**: Gather new or loyal customers through a low cost channel
- **Publisher**: Guarantee a stable and predictable revenue source
- **Platform**: Maximize overall welfare (& utility) with user satisfaction
Effective Cost Per Mille (eCPM)

Expected/estimated revenue per (a thousand) impression(s)
eCPM = Bid Amount x Likeness

Revenue:
Amount that an advertiser is willing to pay for desired actions

Relevance:
How much does a user like to do the actions
In CPC,

eCPM = \frac{BA_{\text{click}} \times \ pCTR}{\text{number of clicks} / \text{number of impressions}}
Leave $(y = 0)$ 

Click $(y = 1)$

$\text{Ad} \perp X$

$X$: Traffic properties (ADxUSRxPLx…)

if(kakao)
Pr\(y = 1 \mid x, \text{ad}\)

Reactive method vs Predictive method

Aggregation of historical data

Learning from historical data
More likely to click

Less likely to click

Logistic Regression (Maximum entropy)

Softmax of binary (1/0) output

\[ \Pr(y = 1 \mid x) = \frac{1}{1 + \exp(-w^T x)} \]
\[ \Pr(y = 1 \mid x) = \frac{1}{1 + \exp(-w^T x)} \]

\[ \text{Loss} = \|y - \hat{y}\|_2 \]
Find $\mathbf{w}$ that minimizes the negative log likelihood (w/ L$_2$ regularization)

$$\arg \min_{\mathbf{w}} \sum_{i=1}^{n} \log(1 + \exp(-y_i\mathbf{w}^T\mathbf{x}_i)) + \frac{\lambda}{2} \left\| \mathbf{w} \right\|_2^2$$

NLL for logistic regression

Control model complexity
\[ w_{t+1} = w_t - \eta_t g_t \]

\[ \eta_t = \frac{\alpha}{\beta + \sqrt{\sum_{s=1}^{t} g_s^2}} \]
FTRL-Proximal (Online)

\[ w_{t+1} = \arg \min_w \left( g_{1:t} \cdot w + \frac{1}{2} \sum_{s=1}^{t} \sigma_s \| w - w_s \|_2^2 + \lambda_1 \| w \|_1 \right) \]

Follow-the-leaders

Proximal (convexity for stability)

Regularization (sparsity)

\[ w_{t+1} = w_t - \eta_t g_t \]

Reference: Ad click prediction: a view from the trenches (2013)
Recent advances on response prediction

- FM/FFM/FwFM (**Factorization Machines**)
- Deep & Cross Network (DCN)
- Model ensemble

\[
\text{Interaction & latent} \quad \Pr(y=1|X) = \frac{1}{1 + \exp(-y*FM(X))}
\]

\[
\text{Nonlinear embedding}
\]

\[
\text{New state-of-the-art}
\]
Pr(y = 1 | x) = 1 / (1 + \exp(-w^T x))

** The pictograms are only for explanation, in that it does not imply that Kakao uses such audience information.
(Hopefully) Almost activities in Kakao ($+\alpha$) with cautious treatments

- Law and guidance (i.e., Privacy)
- De-identification and k-anonymity
- Abstraction: Estimation and aggregation
- No merging between internal and external user data
- & technical and economical barriers
<table>
<thead>
<tr>
<th>M</th>
<th>F</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>ADF</th>
<th>SUBS</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Estimation**

**Aggregation**

**Classification** (Bayesian inference) → Gender/Age-band estimation

**Clustering / LDA** (topic modeling)

**Hashing trick** → Simple but syntactic

**Gradient Boosting Tree** → Effective but slow

**DNN / AE** → Promising as others do
$P(y=1|X) = \sigma(W^T X)$
Conversion really matters.

\[ eCPM = BA_{\text{conv.}} \times pCTR \times pCVR \]

Very similar to CTR but totally different from CTR
Impression ➔ Click ➔ Conversion

- CTR ➔ CVR
- \( \Delta t \) ➔ \( \Delta t \)

Convertion delay: Installation, registration, purchase, subscription, ...

Segment: Personal, Granularity: Variety

One action: Sequence of actions

SSP/DSP: MAT/SDK/Pixel

Data Integrity: Context (hurdles)

Rare event

if (kakao)
Features over Algorithms

- Conversion proxy: Retargeting
- Conversion-driven LookALike
More topics for better response prediction

— Multi-task learning (for each ad, objective)
— Transfer learning
— Landscape forecasting
— Multi-touch attribution
— Cold-start
  — Exploitation vs Exploration
  — Thompson sampling
eCPM = BA x pCTR

How much to bid?
How to dynamically adjust bids?
Is every audience equally valuable to me?
Budget smoothing and **auto-bidding**

- Value/Bidding line (1st price)
- BA*
- Winning curve (2nd price)

**Over-paid for less valuable traffic** vs **Failure in bidding (Under-spent)**
<table>
<thead>
<tr>
<th>Creative</th>
<th>$BA_{clk}$</th>
<th>pCTR</th>
<th>eCPM</th>
<th>Rank</th>
<th>*PPC (GSP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1,000</td>
<td>17%</td>
<td>170</td>
<td>1</td>
<td>941</td>
</tr>
<tr>
<td>B</td>
<td>1,500</td>
<td>6%</td>
<td>90</td>
<td>4</td>
<td>**500</td>
</tr>
<tr>
<td>C</td>
<td>1,200</td>
<td>9%</td>
<td>108</td>
<td>3</td>
<td>1,000</td>
</tr>
<tr>
<td>D</td>
<td>800</td>
<td>20%</td>
<td>160</td>
<td>2</td>
<td>540</td>
</tr>
</tbody>
</table>

* PPC = $BA \times \frac{\text{next eCPM}}{\text{own eCPM}} = \frac{\text{next eCPM}}{\text{own pCTR}}$
** Reserve price = 500
Problem & ideation

- Model validity
- Effect simulation

Validity & revenue
- A/A Test
- A/B/C/… Test
- Random bucket

Complexity & Stability
In theory (model validation)
- Loss (logLoss)
- RIG = 1 - NE (Entropy)
- Calibration = predicted / actual
- AUC

In reality
- **Revenue** per request
More topics beyond this presentation

- Yield optimization (SSP)
  - Auction design (GSP/VCG, reserve price/bid floor)
- Fraud/abusing detection
- Targeting/retargeting/LookALike
- Frequency/recency capping
- AdBlock & DNT: Usefulness vs Annoying
- Dynamic/personalized creative generation
- System consideration (e.g., distributed system)
- Knowledge representation beyond audience
Bid amount (BA) \times \text{Response Rate} = \text{eCPM}

- **Manual setting (by Adv.)**
- **Auto-bidding (BA*)**
- Impressions (\equiv \text{requests})
- Viewable impression
- Clickthrough rate (CTR)
- Conversion rate (CR,

Landscape forecasting (ARIMA, Prophet)

Probability (Logistic Regression)
- Input feature (X)
- Model weight (W)
<table>
<thead>
<tr>
<th><strong>Ranking</strong></th>
<th>eCPM = BA_{imp} = BA_{clk} \times pCTR = BA_{conv} \times pCTR \times pCVR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data/Feature</strong></td>
<td><strong>Privacy</strong>-free audience data</td>
</tr>
<tr>
<td><strong>Embedding</strong></td>
<td>Classification (Bayesian) + Topic Modeling (<strong>LDA</strong>)</td>
</tr>
<tr>
<td><strong>Prediction</strong></td>
<td>Logistic Regression</td>
</tr>
<tr>
<td><strong>Training</strong></td>
<td>(Online) FTRL-Proximal</td>
</tr>
<tr>
<td><strong>BidAmount</strong></td>
<td>Auto-bidding</td>
</tr>
<tr>
<td><strong>Targeting</strong></td>
<td>Conversion-driven LookALike</td>
</tr>
</tbody>
</table>
Ideality vs Reality

Hard works from theory to production

Privacy is our top priority.
Some references
- Ad click prediction: a view from the trenches
- Practical lessons from predicting clicks on ads at facebook
- Simple and scalable response prediction for display advertising
- Modeling delayed feedback in display advertising
- Latent dirichlet allocation
- Factorization machines
- Field-aware factorization machines for CTR prediction
- Field-weighted factorization machines for click-through rate prediction in display advertising
- Deep and cross network for ad click predictions
- Model ensemble for click prediction in bing search ads
- Optimal real-time bidding for display advertising
- Feature hashing for large scale multitask learning
- Score lookalike audiences