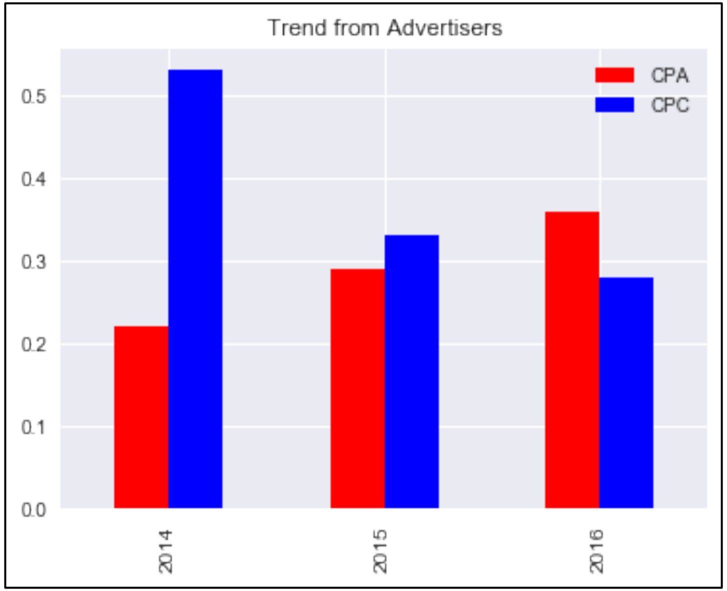
A Practical Framework of Conversion Rate Prediction for Online Display Advertising

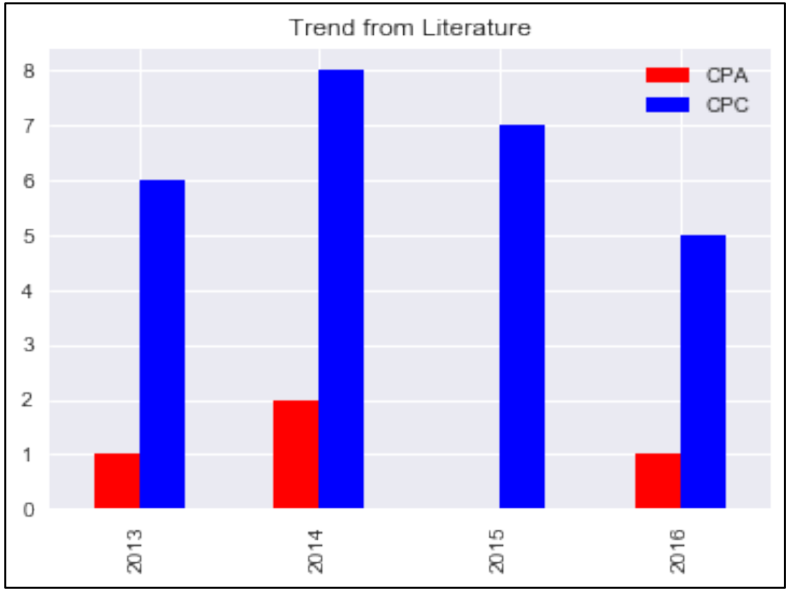
Quan Lu, Shengjun Pan, Junwei Pan, Fengdan Wan

Liang Wang, Hongxia Yang
Alibaba Group.

CPA vs CPC Trend



Percentage of campaigns with each goal type



Number of publications at each year

Challenges for CPA Predictions

Rarity

Extremely low conversion rate. Typically in 10E-5 to 10E-6 range.

Attribution

There is no 100% causal relationship between the events and following conversions

Heterogeneity

Learning across different advertisers is not allowed.

Delayed Feedback

Days or weeks delay before getting the response.

Different types

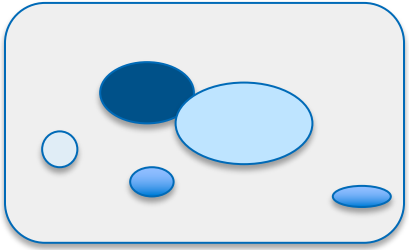
Post-click-action and post-view-action are naturally different in modeling.

Challenges of Conversion Predictions Due to Conversion Rarity

- Worse performance
- Low delivery
- Costly exploration

CVR Safe Prediction Framework

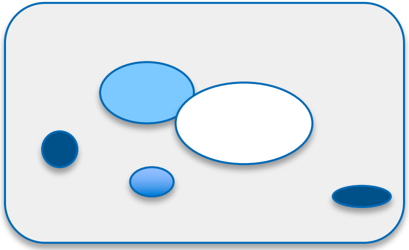
1



Initial stage

Exploration regions are derived from the trained GBDT models. Each region may have different CVR priors.

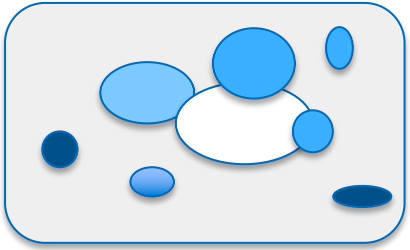
2



Evolution stage

Each region’s predicted CVR is converted to its own empirical mean. So, some regions’ outputs automatically fade out.

3



Growing stage

New exploration regions are added based on the cross region conversions to provide continuously explorations.

Performance Comparison

Campaign	Machine-Learning Tree			Machine-Learning + Data-Driven Tree			CVR lift	eCPA drop	imps lift
	CVR	eCPA	imps	CVR	eCPA	imps			
Campaign1	6.96E-06	63.5419	1,174,131	7.06E-06	61.9564	1,293,526	1.52%	-2.50%	10.17%
Campaign2	5.38E-06	130.2857	1,300,893	9.66E-06	83.8571	1,448,639	79.60%	-35.64%	11.36%
Campaign3	1.49E-03	0.8359	576,431	2.13E-03	0.5987	678,384	42.78%	-28.37%	17.69%

Table 1: Online test results from 3 campaigns. It shows performance and delivery improvements after adding data-driven trees on top of machine-learning trees.

Conversion Adjustment for Delayed Feedbacks

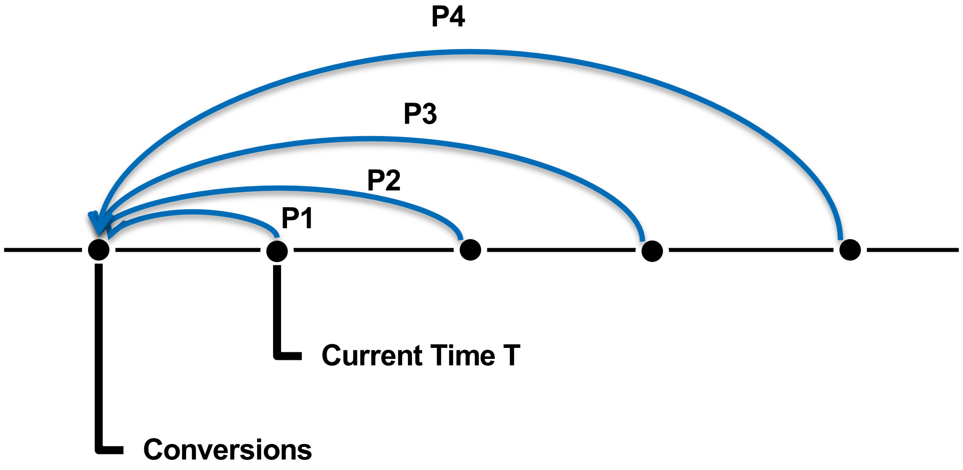
- Trade-off of including new data with compensating empirical estimation.
- Formulate and solve as the following linear programming problem.

$$\min \sum_{j=T+1}^{n} \left[\sum_{i=j-T}^{j} \left(\sum_{j'=i}^{i+T} C_{i,j'} \right) \cdot P_{j-i,D(i)} - \sum_{i=j-T}^{j} C_{i,j} \right]^2$$

$$s.t.$$

$$\sum_{k=0}^{T} P_d(k) = 1 \text{ for } d = 0, 1, \ldots, 6,$$

$$0 \leq P_d(k) \leq 1 \text{ for } k = 0, 1, \ldots, T.$$



Adjusted Conversions at T-1 = N1 / P1

Bid Price Adjustment for Last-Win-All Attribution

- For post-view-conversion campaign, impression value decreases when user has been shown the same impression before.
- The discounted value positively correlates to the elapsed time between the last shown impression at the same user
- The key is to conduct noise reduction and do fast approximation of the integration value online

Algorithm 1 Value Adjustment Factor Estimation

Input: time related parameters:

t_0 : time the previous impression was shown;

t : the current time to bid for a new impression;

T : length of attribution window (in hours);

W : length of conversion window (in hours).

Input: n hours of historical data.

1: For $i, j \in \{1, 2, 3 \ldots, n\}$, find counts $C_{i,j}$, the number of conversions from hour j that are attributed to impressions shown at hour i.

2: Calculate the probability mass of $\kappa(x)$ at hour $h \leq W$:

$$\hat{\kappa}(h) = \frac{\sum^{n}_{\substack{j=T+1 \\ \text{hod}(j)=\text{hod}(h)}} \sum_{i=1}^{j} C_{i,j}}{\sum_{j=T+1}^{n} \sum_{i=1}^{j} C_{i,j}} \Big/ \left|\{h' \leq W : \text{hod}(h') = \text{hod}(h)\}\right|,$$

and we set $\hat{\kappa}(h) = 0$ for $h > W$. Note that we excluded the first T hours of conversions since their attributed impressions may be incomplete.

3: Calculate the probability mass of $q(x - t_0)$ at hour h:

$$\hat{q}(h - t_0) = \frac{\sum^{n}_{\substack{j=T+1 \\ j>h-t_0}} C_{j-(h-t_0),j}}{\sum_{j=T+1}^{n} \sum_{i=1}^{j} C_{i,j}}.$$

4: Calculate attribution probabilities:

$$\hat{f}_{t_0}(t) = \sum_{h=t}^{t_0+T} \hat{q}(h - t_0) \cdot \hat{\kappa}(h), \text{ and } \hat{f}_t(t) = \sum_{h=t}^{t+T} \hat{q}(h - t_0) \cdot \hat{\kappa}(h).$$

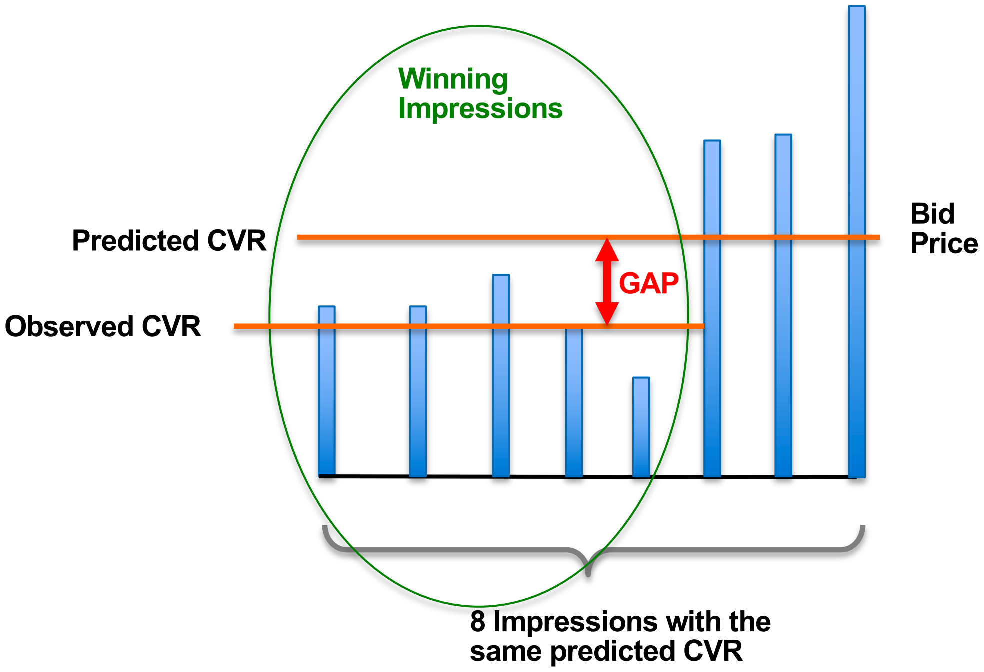
Output: estimated value adjustment factor

$$1 - \frac{\hat{f}_{t_0}(t)}{\hat{f}_t(t)} = 1 - \frac{\sum_{h=t}^{t_0+T} \hat{q}(h - t_0) \cdot \hat{\kappa}(h)}{\sum_{h=t}^{t+T} \hat{q}(h - t_0) \cdot \hat{\kappa}(h)}.$$

Over-prediction Gap Brought by RTB

- Gaps between observations and predictions could be partially introduced by real-time bidding.

- For CPA remarketing impressions, this phenomenon is more obvious due to sharing common information



Problems are not stop signs, they are guidelines

- Robert H. Schuller