Dual Learning Algorithm for Delayed Feedback in Display Advertising

Yuta Saito[1], Gota Morishita[2], Shota Yasui[2]

1.Tokyo Institute of Technology 2.CyberAgent, Inc.

Tokyo Tech CyberAgent AI Lab

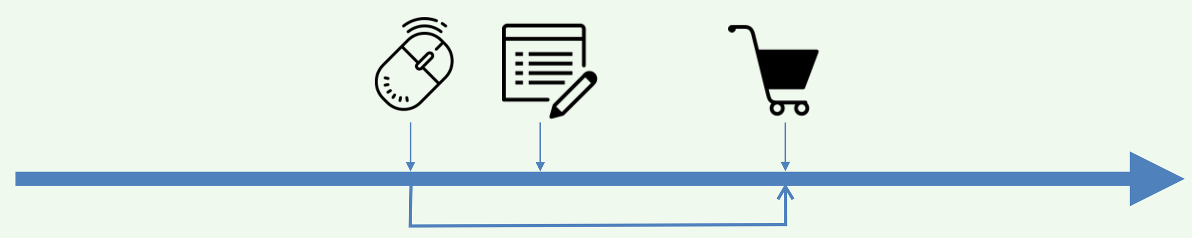
Overview

In display ad, it is important to predict a conversion rate (CVR) from the log data. However, there is a delay between a click and its conversion. Because of the delay, two major difficulties arise.

- **Some positive samples are not correctly observed.**
- **The delay mechanism is not uniform among samples.**

They cause a sever bias in the *naive* empirical loss function. To solve the two problems **simultaneously**, we propose two unbiased estimators: one for the conversion rate and the other for the bias estimation.

Subsequently, we propose an interactive learning algorithm, *Dual Learning Algorithm for Delayed Feedback (DLA-DF).*



Problem Setting and Notation

$X_i \in \mathcal{X}$: features, Y_i: true label, Y_i^{obs} : observed label. Due to the delayed feedback, $Y_i^{obs} \neq Y_i$ for some samples, especially ones collected just before the training begins.

Let O_i be an indicator of being correctly observed,

$$Y_i^{obs} = Y_i \cdot O_i .$$

Define $\theta(X) = P(O = 1 \mid X)$, $\gamma(X) = P(Y = 1 \mid X)$.

The goal is to obtain a hypothesis $f : \mathcal{X} \rightarrow (0,1)$ that predicts $\gamma(\cdot)$. To this end, we want to minimize the ideal loss function $\mathcal{L}_{cvr}(f)$:

$$\mathcal{L}_{CVR}(f) = \mathrm{E}_{X,Y}[Y\delta^{(1)}(f) + (1-Y)\delta^{(0)}(f)] .$$

However, Y_i is not observable, and thus its empirical loss $\widehat{\mathcal{L}}_{cvr}(f)$ cannot be computed from the observed data $\{X_i, Y_i^{obs}\}$. Using Y_i^{obs} to compute the loss instead of Y_i introduces bias.

The critical component of the delayed feedback problem is to estimate the ideal loss function using the observed data.

Proposed Estimators

- **Unbiased Conversion Rate Prediction**

We use $\theta(X)$ as the propensity score for the delayed feedback. However, O_i is not observed. Hence, combining the estimation technique in the field of positive-unlabeled (PU) learning, we define the unbiased IPS estimator:

$$\widehat{\mathcal{L}}_{IPS}(f) = \frac{1}{N}\sum_{i=1}^{N} Y_i^{obs}\left(\frac{1}{\theta(X_i)}\delta_i^{(1)}(f) + \left(1-\frac{1}{\theta(X_i)}\right)\delta_i^{(0)}(f)\right) + (1-Y_i^{obs})\delta_i^{(0)}(f) .$$

$$\mathbb{E}[\widehat{\mathcal{L}}_{IPS}(f)] = \mathcal{L}_{CVR}(f) .$$

- **Unbiased Propensity Estimation**

The unbiasedness stated above depends on the availability of the true propensity score. However, we cannot observe O_i.

Therefore, we propose an empirical loss function that is unbiased against the ideal loss function for the propensity estimation.

$$\widehat{\mathcal{L}}_{ICVR}(g) = \frac{1}{N}\sum_{i=1}^{N} Y_i^{obs}\left(\frac{1}{\gamma(X_i)}\delta_i^{(1)}(g) + \left(1-\frac{1}{\gamma(X_i)}\right)\delta_i^{(0)}(g)\right) + (1-Y_i^{obs})\delta_i^{(0)}(g) .$$

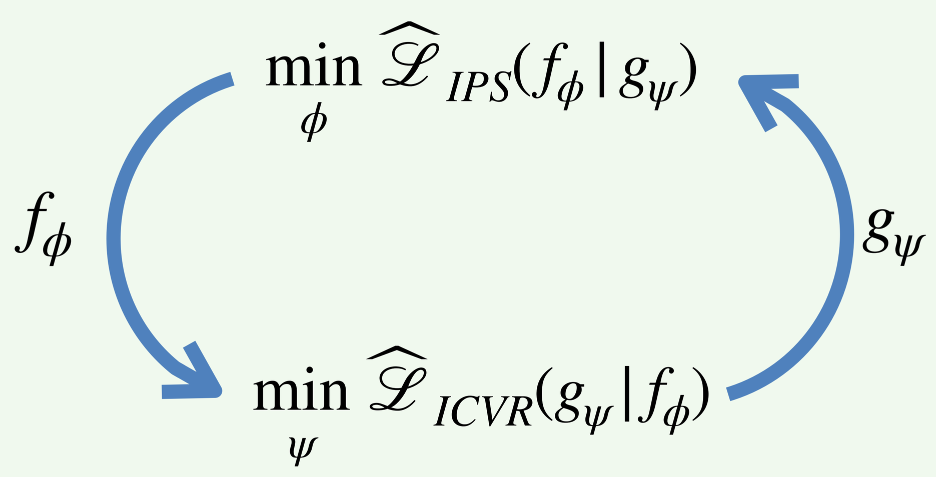
$$\mathbb{E}[\widehat{\mathcal{L}}_{ICVR}(g)] = \mathcal{L}_{Score}(g) .$$

Proposed Learning Algorithm

The unbiasedness of the estimators depends on each other. We proposed DLA-DF, where the two estimators are alternatively learned using $f_\phi(X_i), g_\psi(X_i)$ instead of $\gamma(X_i), \theta(X_i)$, respectively.

f_ϕ : a conversion predictor parameterized by ϕ.

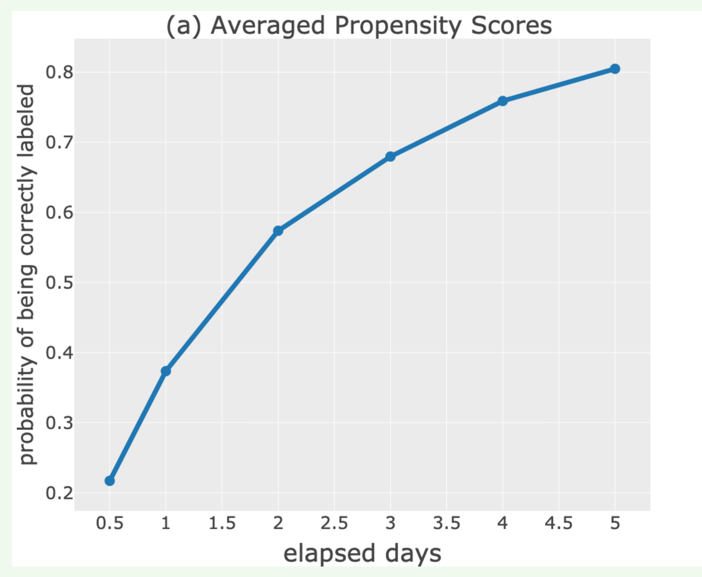
g_ψ : a propensity estimator parameterized by ψ.



Synthetic Experiment

• Data Generation Procedure

1. N click events, p features. $N = 100000, p = 10$.
2. Draw feature vectors $\mathbf{X}$ from $\mathcal{D}_\mathrm{X}$.
3. Decide the training period E.
4. Sample the timestamps of clicks ts_click_i from $U(0,E)$.
5. Sample the delay between clicks and conversions D_i from $Exp(\lambda(X_i))$.
6. Decide the true label Y_i according to whether $\gamma(X_i) > 0.5$.
7. Decide the indicator $O_i = 1\{ts\_click_i + D_i \leq E\}$.
8. Decide the observed label $Y_i^{obs} = O_i \cdot Y_i$.



Averaged propensity score by elapsed days E , that is, the probability of being correctly observed.

When the training period E is shorter, the fewer samples are observed.

• Experiments

We conducted experiments to evaluate ours as well as the existing methods.

- Logistic Regression (LR)
- Non-negative Positive-Unlabeled Learning (nnPU)
- Delayed Feedback Model (DFM)
- Dual Learning Algorithm for Delayed Feedback (nnDLA-DF)

Performance measures:

Relative cross entropy (RCE): the cross entropy of the model (CE_{model}) divided by that of the oracle model (CE_{oracle}) which is trained on the true label Y_i , and thus is the best achievable prediction performance.

$$\mathrm{RCE} = \frac{CE_{model}}{CE_{oracle}} .$$

Results: **Ours performs better when the delay is severe.**