Dual Learning Algorithm for Delayed Conversions

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

Gota Morishita*
Independent Researcher
gota.morishita@gmail.com

Shota Yasui
CyberAgent, Inc.
yasui_shota@cyberagent.co.jp

ABSTRACT

In display advertising, predicting the conversion rate (CVR), meaning the probability that a user takes a predefined action on an advertiser's website, is a fundamental task for estimating the value of displaying an advertisement to a user. There are two main challenges in CVR prediction due to delayed feedback. First, some positive labels are not correctly observed in training data because some conversions do not occur immediately after a click. Second, delay mechanisms are not uniform among instances, meaning some positive feedback are much more frequently observed than others. It is widely acknowledged that these problems lead to severe bias in CVR prediction. To overcome these challenges, we propose two unbiased estimators: one for CVR prediction and the other for bias estimation. Subsequently, we propose a dual learning algorithm in which a CVR predictor and a bias estimator are trained in alternating fashion using only observable conversions. The proposed algorithm is the first of its kind to address the two major challenges in a theoretically sophisticated manner. Empirical evaluations using synthetic datasets demonstrate the practical value of the proposed approach.

CCS CONCEPTS

• **Information systems → Computational advertising**.

KEYWORDS

delayed feedback; dual learning; inverse propensity score

1 INTRODUCTION

Display advertising is a way of online advertising in which advertisers pay publishers for placing ads on their websites. Over the past decade, selling display advertisements via programmatic instantaneous auction called real-time bidding has become a common practice in the display advertising domain [6]. Advertisers are offered several payment options, such as paying per click and

paying per conversion (CPA). CPA has become the predominant payment method because conversions have a more direct effect on advertiser returns on investment compared to clicks. Therefore, we consider a CPA model in which advertisers pay only if a user performs a predefined conversion. A platform that supports such performance-based payment options must convert advertiser bids into an expected price per impression (eCPM) to determine the optimal bid price in an auction [1, 8]. In a CPA model, eCPM depends on the conversion rate (CVR), and accurately predicting the CVR is essential for determining the optimal price to bid for each impression.

Although click-through rate prediction has been extensively studied [2], it is difficult to apply these methods directly to the CVR prediction task. This is because a predictive model should be trained on fresh data to prevent data from becoming stale and to follow seasonal trends [1, 5]; there are two main difficulties in using fresh data for CVR prediction due to delayed feedback issue. First, unlike a click event, a conversion does not always occur immediately after a click on an ad. While the time delay between an impression and click is usually only a few seconds, the time gap between a click and conversion can be a few hours or even days. Consequently, some conversions that will occur eventually have not yet been observed at the time of model training, and the corresponding instances are falsely considered as negative responses (**Positive-Unlabeled problem**). The second challenge is that the missing mechanism for conversion data is missing-not-at-random (MNAR). For example, decisive users are much more likely to convert immediately after a click than indecisive users. Therefore, the probabilities of conversions being observed correctly are not uniform among samples. It is widely recognized that the MNAR mechanism can lead to sub-optimal and biased estimations (**MNAR problem**) [3, 7].

Several works have been conducted to address the delayed feedback issue. [1] assumed that the delay distribution is exponential and proposed two models for predicting the CVR and delay distribution separately. However, this parametric assumption is often too strict for modeling complex real-world conversion data [5, 8]. [8] extended this study and proposed a non-parametric kernel density model for the estimation of delay distributions. However, this kernel method is considered to be unsuitable for the high-dimensional computational advertising domain because of the curse of dimensionality. [5] introduced variants of the importance weighting estimator and positive-unlabeled learning as two separate approaches to solving the delayed feedback problem. However, importance weighting only tackles the MNAR problem and positive-unlabeled learning only addresses the positive-unlabeled problem. As discussed above, one has to address both the positive-unlabeled and MNAR problems for handling delayed feedback, but a method that simultaneously solves these two challenging problems has not yet been proposed.

To address the two major challenges, we first propose an unbiased estimator for the ideal loss function for CVR prediction. The proposed estimator weights each observed conversion using a parameter called the propensity score and does not make any parametric assumptions regarding the delay distribution. However, there is a difficulty in that true propensity scores are unknown in the real-world, thus they have to be estimated. To estimate propensity scores accurately, we subsequently show that the unbiased propensity estimation is possible by weighing each sample using their CVR. Based on these observations, we propose a *Dual Learning Algorithm for Delayed Feedback (DLA-DF)*, which trains a CVR predictor and propensity score estimator in alternating fashion. The proposed learning framework can solve the positive-unlabeled and MNAR problems simultaneously and is expected to adjust to real-world complex delay distributions.

Finally, to evaluate the efficacy of the proposed approach in a delayed feedback setting, we conducted experiments using synthetic datasets. The results demonstrate that the proposed algorithm outperforms existing baselines, particularly in situations where delay is severe and the parametric assumptions of previously proposed methods are violated. These theoretical and empirical findings suggest that the proposed learning framework is a suitable choice for predicting CVR in realistic delayed feedback environments.

2 PROBLEM SETTING

Given a set of N units indexed by i, $X_i \in \mathcal{X} \subset \mathbb{R}^d$ denotes the feature vector for each unit. Let $Y_i \in \mathcal{Y} = \{0, 1\}$ be a random variable representing true conversion information. If an individual i will eventually convert, then $Y_i = 1$. Otherwise, $Y_i = 0$. In the delayed feedback setting, true conversion variables are not fully observable because of conversion delay. To formulate such a delayed feedback setting precisely, we introduce another binary random variable $O_i \in \{0, 1\}$. This random variable represents whether or not a true outcome is observed, which depends on the elapsed time from a corresponding click. If $O_i = 1$, then a conversion is observed. Otherwise, a conversion is not correctly observed. Using these random variables, we can represent an observed outcome indicator as $Y^{obs} = O_i \cdot Y_i$. If we have observed the conversion of i, then $Y_i^{obs} = 1$. Otherwise, $Y_i^{obs} = 0$. Note that the true conversion indicator Y_i is not always equal to the observed conversion indicator Y_i^{obs}; the conversion of i is observable only when the unit will eventually convert and the true outcome is observable (i.e., $Y_i^{obs} = 1 \Leftrightarrow O_i = 1 \& Y_i = 1$). Finally, we use $E \in \mathbb{R}_{\geq 0}$ to denote the elapsed time since a click. When E is large, the probability of a true label being correctly observed is also large.

Throughout this paper, we assume that features affecting both O and Y are fully observed (i.e., $Y \perp O \mid X, E$), which is referred to as *Unconfoundedness* in causal inference [3]. Building on this assumption, we obtain the following equation connecting the true CVR to the observed CVR:

$$P(Y_i^{obs} = 1 \mid X_i, E_i) = \theta(X_i, E_i) \cdot \gamma(X_i)$$

where we denote $P(O = 1 \mid X, E)$ as $\theta(X, E)$ and $P(Y = 1 \mid X)$ as $\gamma(X)$. Additionally, the CVR is assumed to be independent of the elapsed time, as described in Eq. (4) in [1].

The goal of this study is to obtain a predictor $f : \mathcal{X} \rightarrow (0, 1)$ that accurately predicts the true CVR. To achieve this goal, we define the ideal loss function that should be optimized to obtain an accurate predictor as follows:

$$\mathcal{L}_{ideal}^{CVR}(f) = \mathbb{E}_{(X,Y)} \left[Y\delta^{(1)}(f(X)) + (1 - Y)\delta^{(0)}(f(X)) \right] \quad (1)$$

where the functions $\delta^{(1)}(\cdot)$ and $\delta^{(0)}(\cdot)$ characterize the loss function. For example, when these functions are defined as $\delta^{(1)}(f) = -\log(f(X))$, $\delta^{(0)}(f) = -\log(1 - f(X))$, Eq. (1) is called binary cross entropy loss.

The loss function in Eq. (1) is defined using the true conversion indicator, and thus, is ideal. However, in the delayed feedback setting, true conversion indicators (Y) are unobserved and the direct minimization of this ideal loss function is infeasible. Therefore, the critical component of the delayed feedback problem is the estimation of the ideal loss function from observable variables.

3 PROPOSED METHOD

3.1 Unbiased CVR Prediction

To approximate the ideal loss function from observable data, here we propose an unbiased estimator for the ideal loss function for CVR prediction.

Definition 3.1. (IPS estimator for the ideal loss function for CVR prediction) *When the set of propensity scores is given, the inverse propensity score (IPS) estimator for the ideal loss function is defined as*

$$\widehat{\mathcal{L}}_{IPS}^{CVR}(f|\theta) = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i^{obs}}{\theta(X_i, E_i)} \delta_i^{(1)}(f) + (1 - \frac{Y_i^{obs}}{\theta(X_i, E_i)})\delta_i^{(0)}(f),$$

$$(2)$$

where $\theta(X, E) = P(O = 1 \mid X, E) = P(Y^{obs} = 1 \mid Y = 1, X, E)$ *is called the propensity score and* $\delta_i^{(\cdot)}(f)$ *is a simplified notation for* $\delta^{(\cdot)}(f(X_i))$.

The following proposition formally proves that the IPS estimator is statistically unbiased against the ideal loss function.

PROPOSITION 3.2. *(Unbiasedness of the IPS estimator) The IPS estimator in Eq. (2) is statistically unbiased against the ideal loss function in Eq. (1), i.e.,* $\mathbb{E}[\widehat{\mathcal{L}}_{IPS}^{CVR}(f)] = \mathcal{L}_{ideal}^{CVR}(f)$.

PROOF. We can prove the unbiasedness by following the same logic flow used in Proposition 4.3 in [7]. □

Proposition 3.2 validates that unbiased CVR prediction is possible by optimizing the unbiased loss function in Eq. (2) using only observable conversions.

3.2 Unbiased Propensity Estimation

The unbiasedness stated in Proposition 3.2 is desirable for obtaining a CVR predictor, but is dependent on the availability of true propensity scores. In general, the estimation of propensity scores in the IPS estimator can be formulated as a classification problem. However, observation indicators are unobservable in our setting. Therefore, we propose a method for the unbiased estimation of propensity scores from observed conversions.

We first define the ideal loss function for propensity estimation as follows:

$$\mathcal{L}_{ideal}^{Score}(g) = \mathbb{E}_{(X,E,O)} \left[O\delta^{(1)}(g(X,E)) + (1-O)\delta^{(0)}(g(X,E)) \right] \tag{3}$$

where $g : \mathcal{X} \times \mathbb{R} \to (0,1)$ is a predictor that estimates the propensity score[1].

We now propose an *inverse conversion rate* (ICVR) estimator that shares the same structure as the IPS estimator.

Definition 3.3. (ICVR estimator) *When a set of CVRs is given, the ICVR estimator for the ideal loss function in Eq. (3) is defined as*

$$\widehat{\mathcal{L}}_{ICVR}^{Score}(g|\gamma) = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i^{obs}}{\gamma(X_i)} \delta_i^{(1)}(g) + (1 - \frac{Y_i^{obs}}{\gamma(X_i)}) \delta_i^{(0)}(g), \tag{4}$$

where $\delta_i^{(\cdot)}(g)$ is a simplified notation for $\delta^{(\cdot)}(g(X_i, E_i))$.

Following the same logic flow presented in Proposition 3.2, the next proposition proves that the ICVR estimator is statistically unbiased against the ideal loss function for propensity estimation.

PROPOSITION 3.4. *(Unbiasedness of the ICVR estimator) The ICVR estimator in Eq. (4) is statistically unbiased against the ideal loss function in Eq. (3), i.e., $\mathbb{E}[\widehat{\mathcal{L}}_{ICVR}^{Score}(g)] = \mathcal{L}_{ideal}^{Score}(g)$.*

Proposition 3.4 indicates that the unbiased propensity estimation is possible by optimizing the unbiased loss function in Eq. (4) using only observable conversions.

3.3 Algorithm

Here, we describe the proposed DLA-DF algorithm, which jointly trains a propensity estimator and CVR predictor using observable conversions.

First, given a propensity estimator g_ϕ parameterized by ϕ, the loss function for deriving the parameters of a CVR predictor f is defined as

$$\widehat{\mathcal{L}}_{IPS}^{CVR}(f \mid g_\phi) = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i^{obs}}{g_\phi(X_i, E_i)} \delta_i^{(1)}(f) + (1 - \frac{Y_i^{obs}}{g_\phi(X_i, E_i)}) \delta_i^{(0)}(f)$$

Next, given a CVR predictor f_ψ parameterized by ψ, the loss function for deriving the parameters of a propensity estimator g is defined as

$$\widehat{\mathcal{L}}_{ICVR}^{Score}(g \mid f_\psi) = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i^{obs}}{f_\psi(X_i)} \delta_i^{(1)}(g) + (1 - \frac{Y_i^{obs}}{f_\psi(X_i)}) \delta_i^{(0)}(g)$$

The detailed procedure for the proposed DLA-DF algorithm is described in Algorithm 1.

3.4 Variance Reduction Technique

The proposed learning framework is theoretically refined and promising, however, unbiased estimators derived using inverse propensity weighting are widely known to exhibit large variance [7]. Therefore, we analyze the variance of the unbiased estimators and propose a method for addressing this variance issue.

Algorithm 1 Dual Learning Algorithm for Delayed Feedback

Input: training data $\mathcal{D} = \{X_i, E_i, Y_i^{obs}\}_{i=1}^{N}$; mini-batch size m; learning rate η
Output: model parameters ψ and ϕ
1: Initialize parameters with random weights ψ, ϕ
2: **repeat**
3:     Sample mini-batch data $\{X_j, E_j, Y_j^{obs}\}_{j=1}^{m}$ from $\mathcal{D}$
4:     Update ψ by $\nabla_\psi \widehat{\mathcal{L}}_{IPS}^{CVR}(f_\psi \mid g_\phi)$ with a fixed ϕ
5:     Update ϕ by $\nabla_\phi \widehat{\mathcal{L}}_{ICVR}^{Score}(g_\phi \mid f_\psi)$ with a fixed ψ
6: **until** convergence;
7: **return** ψ, ϕ

THEOREM 3.5. *(Variance of the unbiased estimators) Given sets of independent random variables $\{(Y_i^{obs}, O_i, Y_i)\}$, propensity scores $\{\theta(X_i, E_i)\}$, and a CVR predictor f, the variance of the IPS estimator is*

$$\mathbb{V}\left(\widehat{\mathcal{L}}_{IPS}^{CVR}(f)\right) = \frac{1}{N^2} \sum_{i=1}^{N} \gamma(X_i) \left(\frac{1}{\theta(X_i, E_i)} - \gamma(X_i)\right) \left(\delta_i^{(1)}(f) - \delta_i^{(0)}(f)\right)^2.$$

Replacing $\gamma, \theta, \delta_i^{(1)}(f)$, and $\delta_i^{(0)}(f)$ with $\theta, \gamma, \delta_i^{(1)}(g)$, and $\delta_i^{(0)}(g)$ yields the variance of the ICVR estimator.

PROOF. We can derive the variance by following the same logic flow presented in Theorem 4.4 of [7]. □

The variance depends on the inverse of the propensity scores, and thus, can be huge, particularly when severe delay occurs. Therefore, we propose utilizing the following *non-negative estimator* [4] to address the variance problem of the unbiased estimators.

Definition 3.6. (Non-negative estimator) *When propensity scores are given, the non-negative estimator is defined as*

$$\widehat{\mathcal{L}}_{non\text{-}negative}(f) = \frac{1}{N} \sum_{i=1}^{N} \max\{\ell_{IPS}(f), 0\}, \tag{5}$$

where $\ell_{IPS}(f) = \frac{Y_i^{obs}}{\theta(X_i, E_i)} \delta_i^{(1)}(f) + (1 - \frac{Y_i^{obs}}{\theta(X_i, E_i)}) \delta_i^{(0)}(f)$. The non-negative variant for the ICVR estimator can be defined similarly. This non-negative estimator provides a lower variance than the IPS and ICVR estimators at the cost of introducing some bias.

4 SYNTHETIC EXPERIMENT

In this section, we present an empirical comparison of the proposed method to baseline methods using a synthetic dataset.

4.1 Experimental Setup

4.1.1 Synthetic data generation procedure. We created a synthetic dataset simulating a delayed feedback setting. The data generation procedure is presented in Algorithm 2[2]. We set $N = 100,000$, $p = 30$, $\sigma_X = 0.5$, and $\sigma_W = 1.0$. For the delay distribution ($\mathcal{D}_{delay}$), we considered exponential and normal distributions. The lengths of the training period L were set to 0.5, 1, 2, 4 (days). A smaller value of L yields smaller propensities, as indicated in Figure 1 (left).

---

[1]E is unavailable for the test data. However, the propensity score estimator is necessary only for training, and thus, the unavailability of E in the test data is not an issue.

[2]sigmoid($\cdot$) is the sigmoid function, $Unif(\cdot, \cdot)$ is a uniform distribution, and $Bern(\cdot)$ is a Bernoulli distribution.
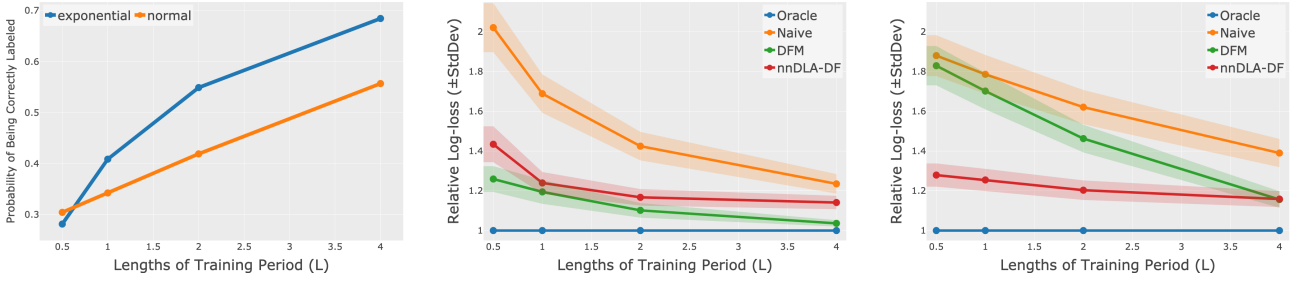
Figure 1: (left) Averaged propensity score for each length of training period. (center) Relative log-loss on test sets when the delay distribution is exponential. (right) Relative log-loss on test sets when the delay distribution is normal.

Algorithm 2 Synthetic data generation process

Input: The number of click events N, number of features observed for each event p, length of training periods L, and standard deviations of the distributions σ_X and σ_W.

Output: dataset $\{X_i, D_i, E_i, Y_i^{obs}\}_{i=1}^n$.

1: Sample true coefficient vectors: $W_{cvr}, W_{expo} \sim \mathcal{N}(\mathbf{0}_p, \sigma_W^2 I_p)$
2: **for** $i = 1, \ldots, n$ **do**
3:    Sample feature vectors: $X_i \sim \mathcal{N}(\mathbf{0}_p, \sigma_X^2 I_p)$.
4:    Calculate true CVRs: $\gamma(X_i) = \text{sigmoid}(W_{cvr} X_i)$
5:    Sample time stamps of click: $ts\_click_i \sim Unif(0, L)$
6:    Sample the lengths of delay: $D_i \sim \mathcal{D}_{delay}(\exp(W_{expo} X_i))$
7:    Calculate the elapsed time from the click: $E_i = L - ts\_click_i$
8:    Calculate the true observation variables: $O_i = \mathbb{I}\{D_i \leq L\}$
9:    Sample the true conversion indicators: $Y_i \sim Bern(\gamma(X_i))$
10:   Calculate the observed conversion indicators: $Y_i^{obs} = O_i \cdot Y_i$
11: **end for**
12: **return** $\{X_i, D_i, E_i, Y_i^{obs}\}_{i=1}^n$

4.1.2 Baselines and the proposed method. We compared the performances of the following methods. **Oracle**: A logistic regression model trained using true conversion data (Y), which is unobservable in the real-world. Therefore, the performance of the oracle model is the best achievable prediction performance. **Naive**: A logistic regression model trained naively using observed conversions (Y^{obs}). **Delayed Feedback Model (DFM) [1]**: This model is a widely used baseline in delayed conversion settings [5, 8] and assumes that the delay distribution is exponential. **Non-negative Dual Learning Algorithm (nnDLA-DF)**: This is the proposed method. We used a logistic regression model for both the CVR predictor (f) and propensity estimator (g). Both estimators were trained using non-negative loss function in Eq. (5).

4.2 Results

Figure 1 (center) and (right) present the values of the log-loss on the test sets relative to the performance of the oracle model when the delay follows exponential and normal distributions, respectively. For both figures, averaged relative log-loss on test sets and its standard deviations over 10 iterations are reported.

Figure 1 (center) demonstrates that the proposed nnDLA-DF is outperformed by DFM. This result is reasonable because the

DFM's assumption of an exponential delay distribution is perfectly satisfied in this setting. However, the proposed method exhibits competitive and stable performance, despite that it does not assume any assumption regarding the delay distribution. Figure 1 (right) demonstrates that the proposed method significantly outperforms the other methods when $L = 0.5, 1, 2$. In contrast, the benefits of DLA-DF are much smaller when $L = 4$, but it is not outperformed by the other methods in any setting, which demonstrates the stable prediction performance of the proposed algorithm.

5 CONCLUSION

In this study, we explored the delayed feedback problem, where true conversion indicators are not fully observable due to conversion delay. To address this problem, we developed the DLA-DF algorithm, which is the first to solve both the positive-unlabeled and MNAR problems of delayed conversions simultaneously. Additionally, the proposed framework does not depend on any parametric assumptions regarding delay distributions and is able to perform well in a wide range of situations. In empirical evaluations, the proposed algorithm outperformed existing baselines, particularly in practical settings where there exists severe delay or the parametric assumptions regarding delay distributions are no longer satisfied.

REFERENCES

[1] Olivier Chapelle. 2014. Modeling delayed feedback in display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1097–1105.
[2] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. ACM, 7–10.
[3] Guido W Imbens and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
[4] Ryuichi Kiryo, Gang Niu, Marthinus C du Plessis, and Masashi Sugiyama. 2017. Positive-unlabeled learning with non-negative risk estimator. In *Advances in neural information processing systems*. 1675–1685.
[5] Sofia Ira Ktena, Alykhan Tejani, Lucas Theis, Pranay Kumar Myana, Deepak Dilipkumar, Ferenc Huszár, Steven Yoo, and Wenzhe Shi. 2019. Addressing delayed feedback for continuous training with neural networks in CTR prediction. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 187–195.
[6] S. Muthukrishnan. 2009. Ad Exchanges: Research Issues. In *WINE*.
[7] Yuta Saito, Suguru Yaginuma, Yuta Nishino, Hayato Sakata, and Kazuhide Nakata. 2020. Unbiased Recommender Learning from Missing-Not-At-Random Implicit Feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 501–509.
[8] Yuya Yoshikawa and Yusaku Imai. 2018. A Nonparametric Delayed Feedback Model for Conversion Rate Prediction. *arXiv preprint arXiv:1802.00255* (2018).