A Feedback Shift Correction in Predicting Conversion Rates under Delayed Feedback

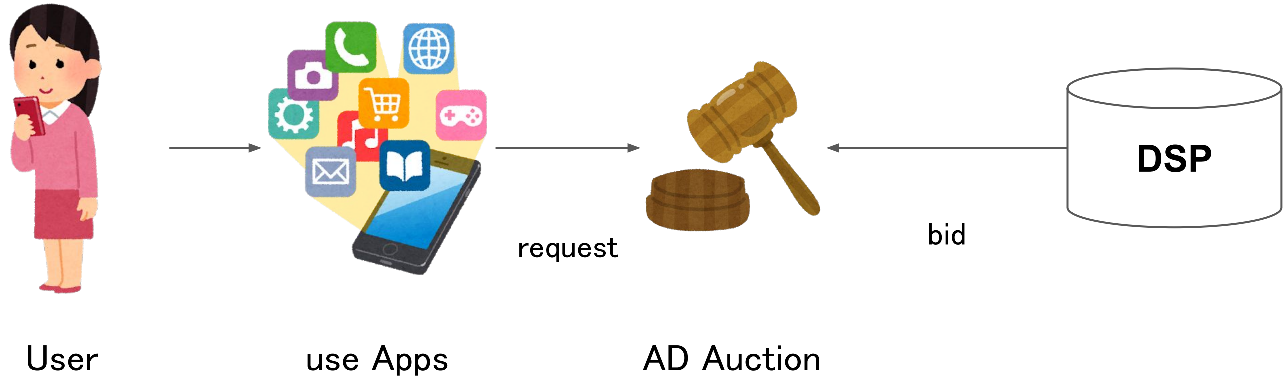
Shota Yasui, Gota Morishita,
Komei Fujita, Masashi Shibata

The Web Conference 2020

Introduction and Problem Setting

Conversion Prediction

Predict Conversion-Rate(CVR) for each request.



Predicting CVR is important to decide the bid price

Ideal loss function

The following loss should be minimized.

$$G \equiv \mathbb{E}_{(x,c)\sim(X,C)}\left[L\left(x, c; \hat{f}(x,\theta)\right)\right]$$

features

Conversion

model

The ideal parameters are as follow

$$\theta^* \in \underset{\theta\in\Theta}{\arg\min}\, G.$$

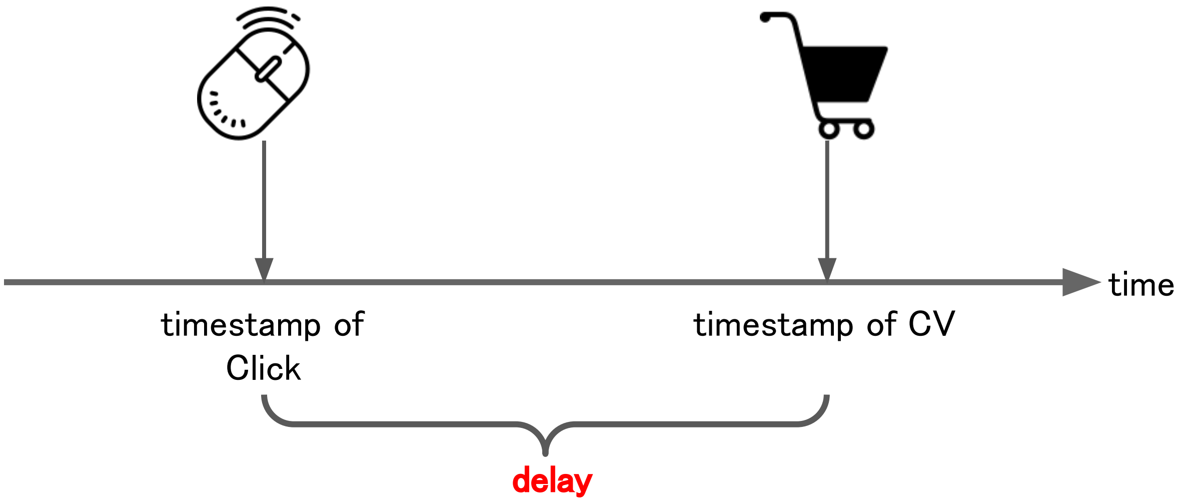
This is not possible!

Because we do not observe c due to the <u>delayed feedback</u>.

Delayed Feedback

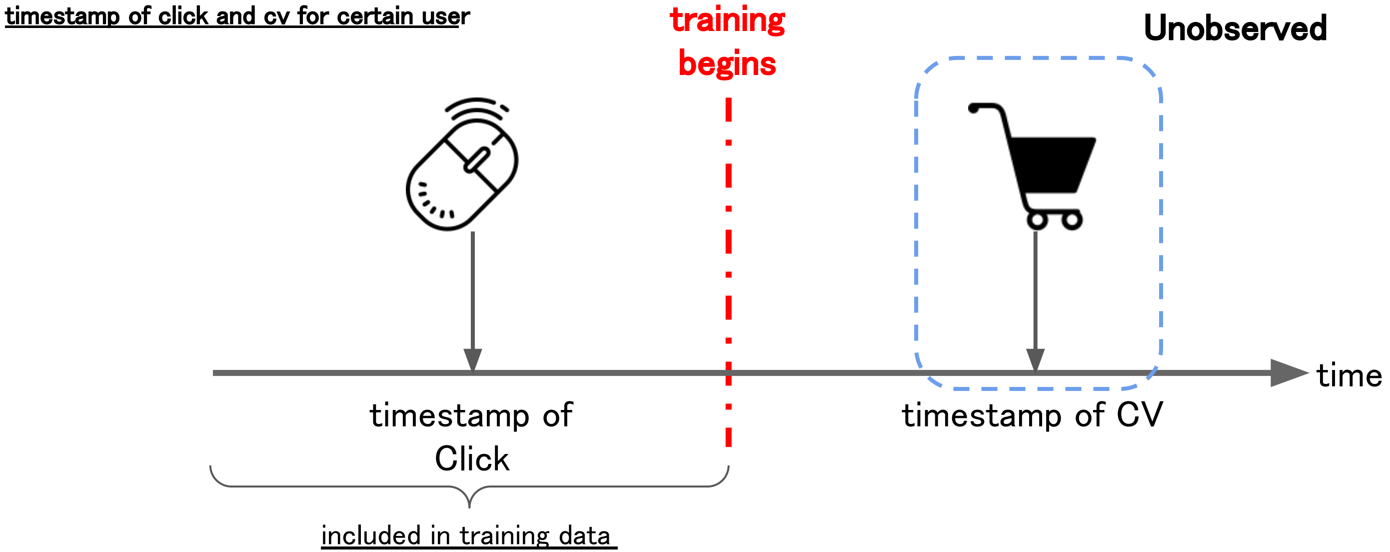
Delayed Feedback

timestamp of click and cv for certain user



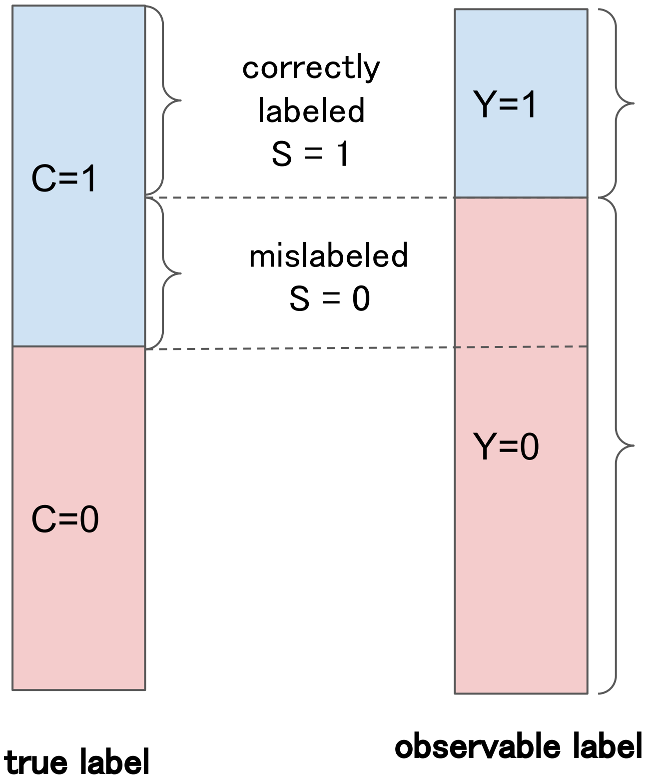
- user takes sometimes to purchase items after clicked the ad.

The problem of Delayed Feedback



- we can not observe CV for this user
- **this sample is recognized as negative label! (mislabeled)**

The relation between Y and C



$$P(Y=1|X=x) = P(C=1|X=x)P(S=1|C=1,X=x)$$

Prob of correctly labeled

$$P(Y=0|X=x) = P(C=0|X=x) + P(S=0,C=1|X=x)$$

Prob of mislabel

Bias in standard supervised approach

$$G \equiv \mathbb{E}_{(x,c)\sim(X,C)}\left[L\left(x, c; \hat{f}(x,\theta)\right)\right]$$

$$\theta^* \in \underset{\theta\in\Theta}{\arg\min}\, G.$$

ideal loss

$$\hat{G}^{(n)} \equiv \frac{1}{n}\sum_{i=1}^{n} L\left(x_i, y_i; \hat{f}(x_i,\theta)\right)$$

$$\hat{\theta}_{ERM} \equiv \underset{\theta\in\Theta}{\arg\min}\, \hat{G}^{(n)}.$$

actual loss(ERM)

$$\mathbb{E}[\hat{\theta}_{ERM}] \neq \theta^*$$

Inconsistent!

Our Solution

Importance Weight Approach

Importance Weight(FSIW) approach

We propose consistent loss based on the Importance Weight(Propensity Score)

<u>**ideal-loss**</u>

$$\mathbb{E}_{(x,c)\sim(X,C)}\left[L\left(x, c; \hat{f}(x,\theta)\right)\right]$$

<u>**Unbiased-loss (consistent?)**</u>

$$\mathbb{E}_{(x,y)\sim(X,Y)}\left[\frac{P(C = y|X = x)}{P(Y = y|X = x)} L\left(x, y; \hat{f}(\theta)\right)\right]$$

Importance Weight

Importance Weight(FSIW) approach

Our empirical loss

$$\hat{G}_{IW}^{(n)} \equiv \frac{1}{n}\sum_{i=1}^{n} \boxed{\frac{P(C = y_i | X = x_i)}{P(Y = y_i | X = x_i)}} L\left(x_i, y_i; \hat{f}(x_i, \theta)\right)$$

Importance Weight

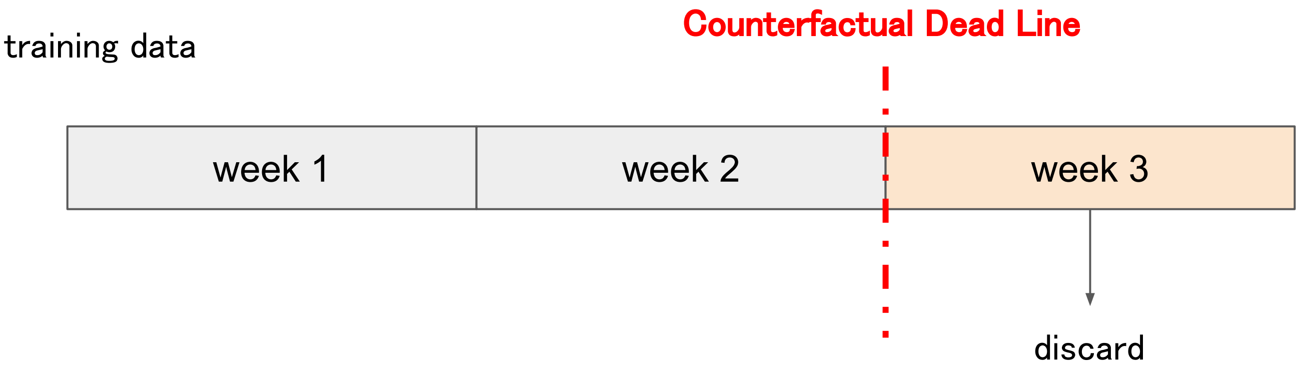
The basic idea is to weight each sample by the conditional density ratio.

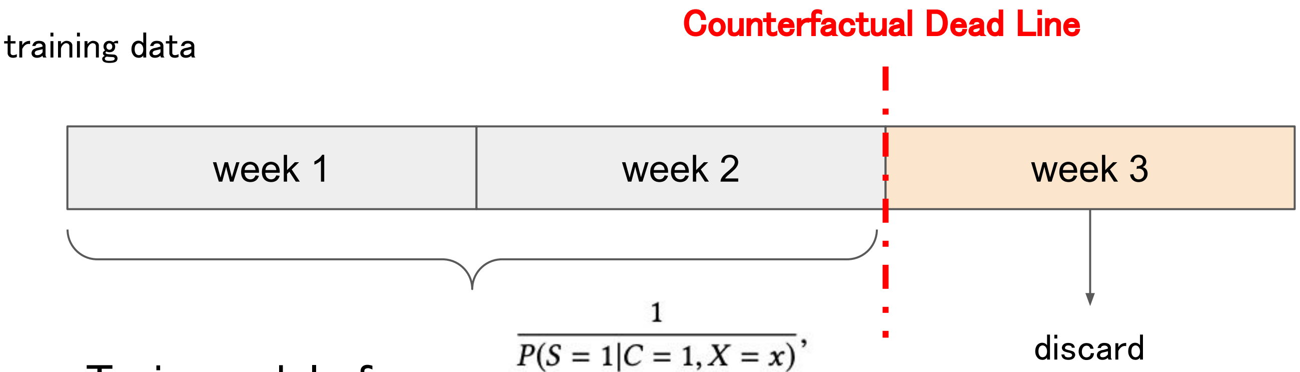
How to estimate FSIW

$$P(Y=1|X=x) = P(C=1|X=x)P(S=1|C=1,X=x)$$
$$P(Y=0|X=x) = P(C=0|X=x) + P(S=0,C=1|X=x)$$

$$\frac{P(C=1|X=x)}{P(Y=1|X=x)} = \frac{1}{P(S=1|C=1,X=x)},$$
$$\frac{P(C=0|X=x)}{P(Y=0|X=x)} = (1 - \frac{P(S=0,C=1|X=x)}{P(Y=0|X=x)}),$$

We estimate these probability from data old enough to observe S and C.


training data
Counterfactual Dead Line
week 1
week 2
week 3
discard



Train models for

$$\frac{1}{P(S=1|C=1,X=x)},$$

$$(1-\frac{P(S=0,C=1|X=x)}{P(Y=0|X=x)}),$$

Counterfactual Dead Line

training data

week 1	week 2	week 3

discard

Train models for

$$\frac{1}{P(S=1|C=1,X=x)},$$

$$(1-\frac{P(S=0,C=1|X=x)}{P(Y=0|X=x)}),$$

training data

Importance weight		
week 1	week 2	week 3

Train the CVR model

features of our proposed method

It is just a importance weight

- can be used for any CVR model
- can fit the delay nonparametrically
- does not increase the time complexity of CVR models

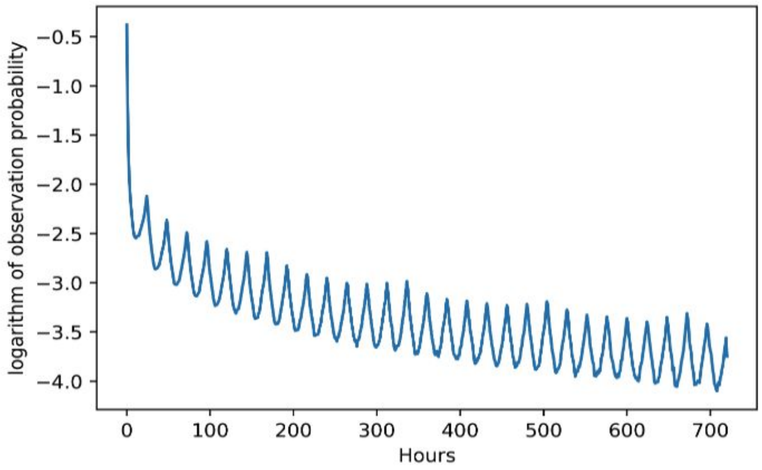
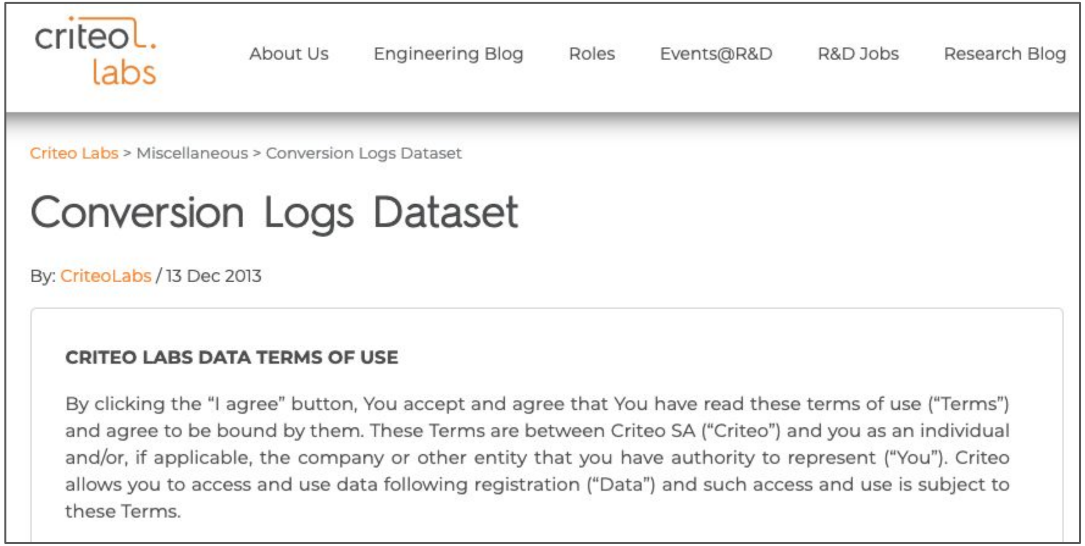


Figure 4: Criteo Dataset: Probability density function of the delays between clicks and conversions. The oscillating shape is a result of daily cyclicality.

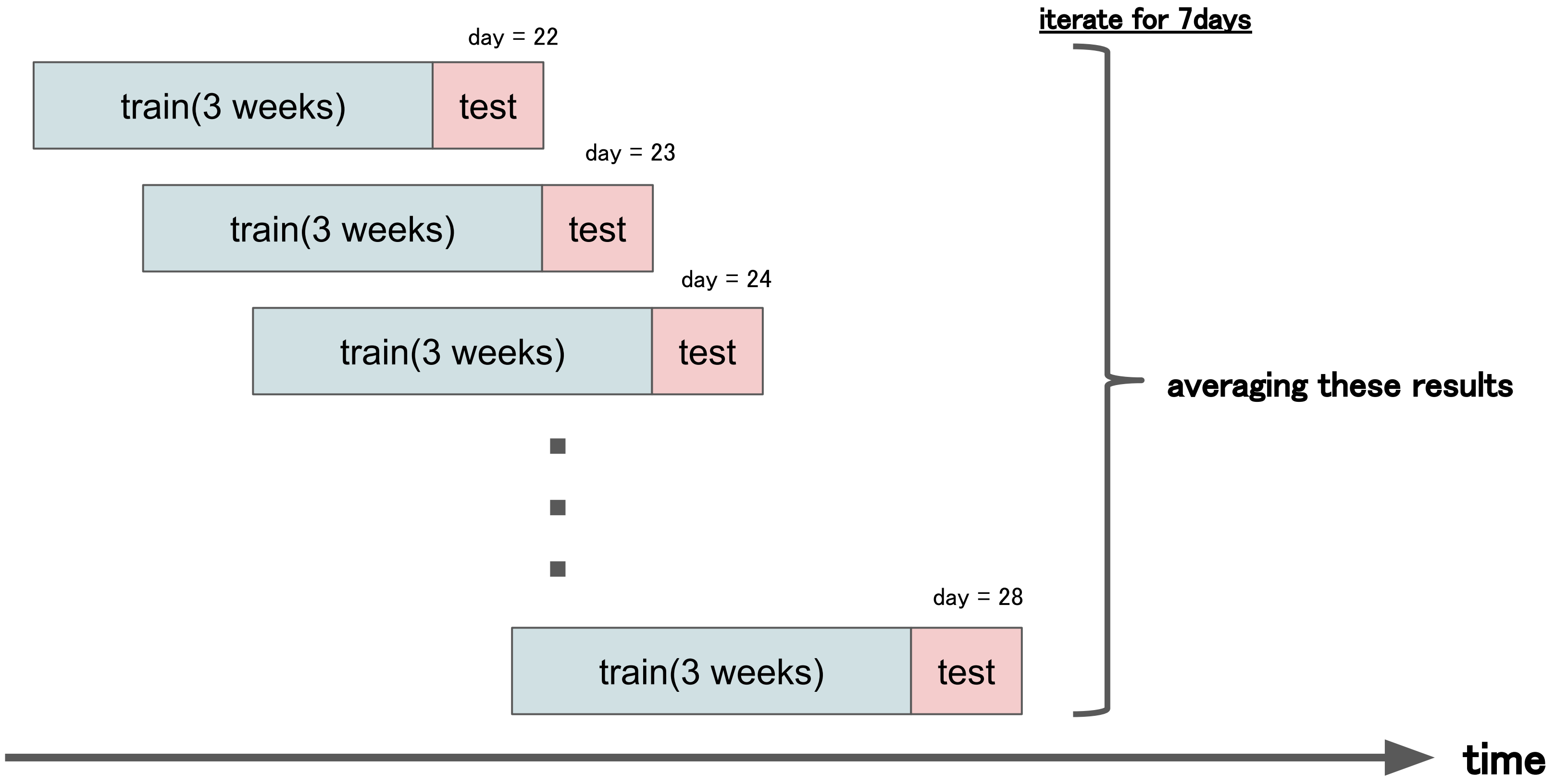
Experiment

Conversion Logs Dataset



- Open data provided by Criteo (Link)
- 30days of click and CV log
- Used in Chapelle(2014)
- observation period is 30days

Experiment procedure



Result 1

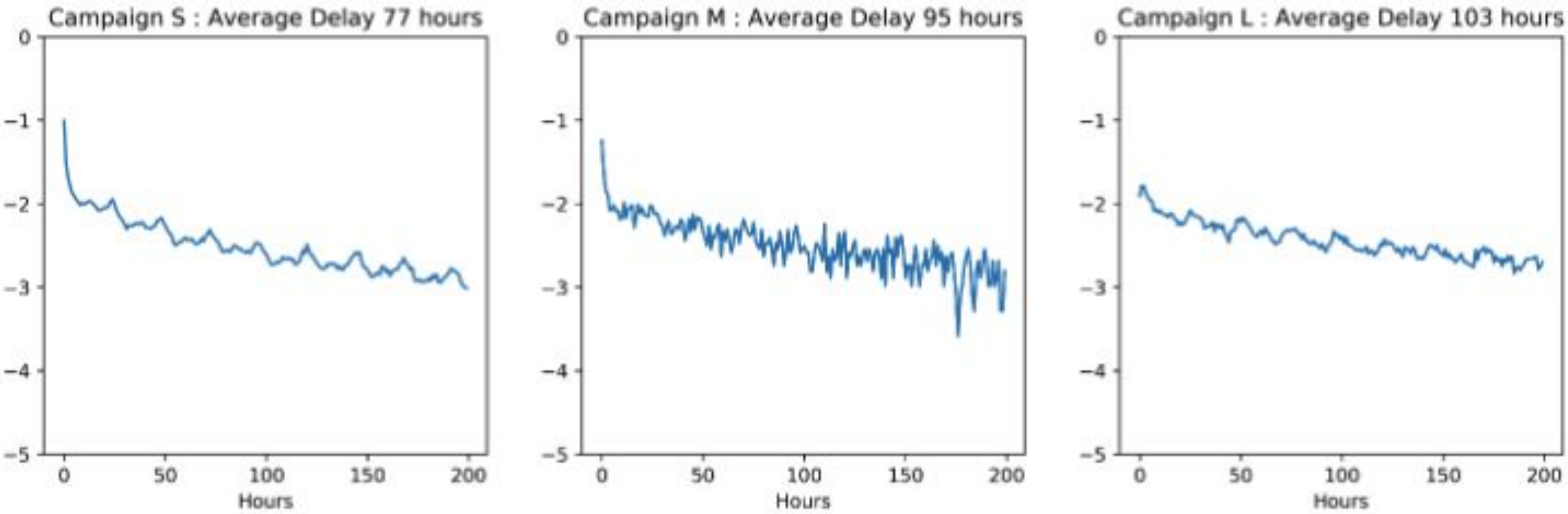
	Pure-Logistic Regression LR	Chapelle(2014) DFM	Proposed Method LR-FSIW
LL	0.4076	0.3989	**0.3928***
PR-AUC	0.6345	0.6481	0.6482
NLL	25.21	27.33	**28.02***

- Normalized-logloss(NLL) is the most important metrics
  - we use prediction probability for bidding
  - logloss(LL) is sensitive to the base CVR

Dynalyst Data

- DSP in Cyberagent.inc
- 2 experiments
    - the same procedure as the first experiment
        - focus on three campaigns
        - baseline model is FFM (Juan 2017)
    - Online A/B test

Three Campaigns



- Observational period is different by campaings
  - S: 1days
  - M: 3days
  - L: 7days

Result 2

		LL	PR-AUC	NLL
Campaign L	FFM	0.3523	0.1612	1.7197
	FFMIW	**0.3500**	**0.1660**	**2.304***
Campaign M	FFM	0.2409	0.0808	0.2160
	FFMIW	**0.2401**	**0.0828**	**0.3771**
Campaign S	FFM	0.4026	0.2055	2.9953
	FFMIW	**0.3967**	**0.2058**	**3.361**

Only Campaign L shows the improvement.

Follow Up Online Experiment@Campaign-L

CV	Cost	CPA
+31%*	+28%*	-2%

Table 5: Online relative comparison of FFM and FFMIW in the conversion(CV), Cost and CPA. The shown values are the relative change in FFMIW against FFM. * means statistical significance.

- Improved cost consumption and CV.
- CPA does not change or slightly decreased.

Conclusion

- **We proposed a consistent loss to predict CVR under Delayed Feedback.**

- **Our method performs better in two offline and one online experiment.**

Thank you for listening!

appendix

cumulative distribution of delay

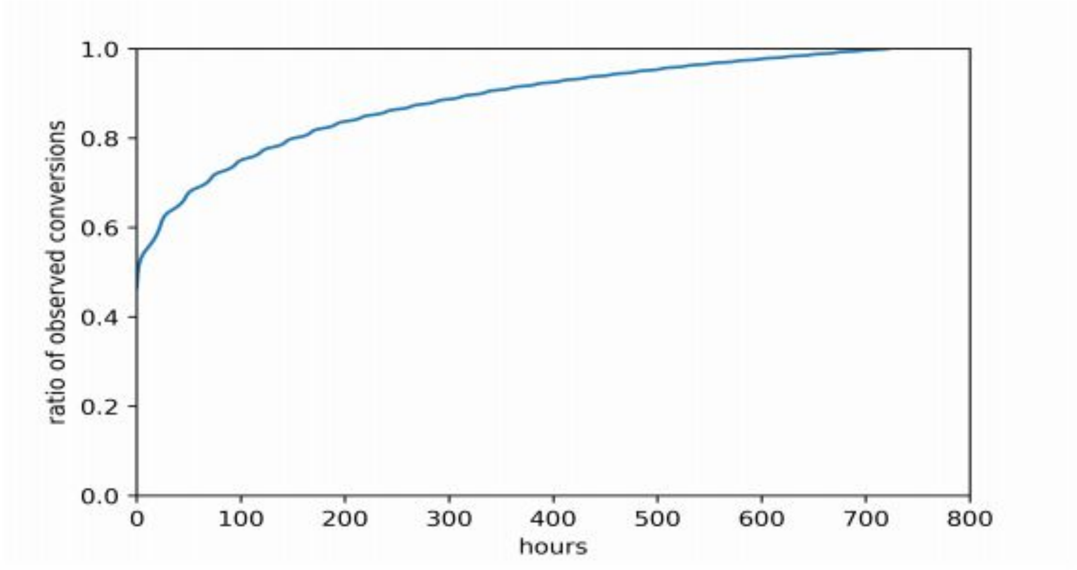


Figure 1: Criteo Dataset: Cumulative distribution of the delay between the click and its conversion.

effect of counterfactual deadline

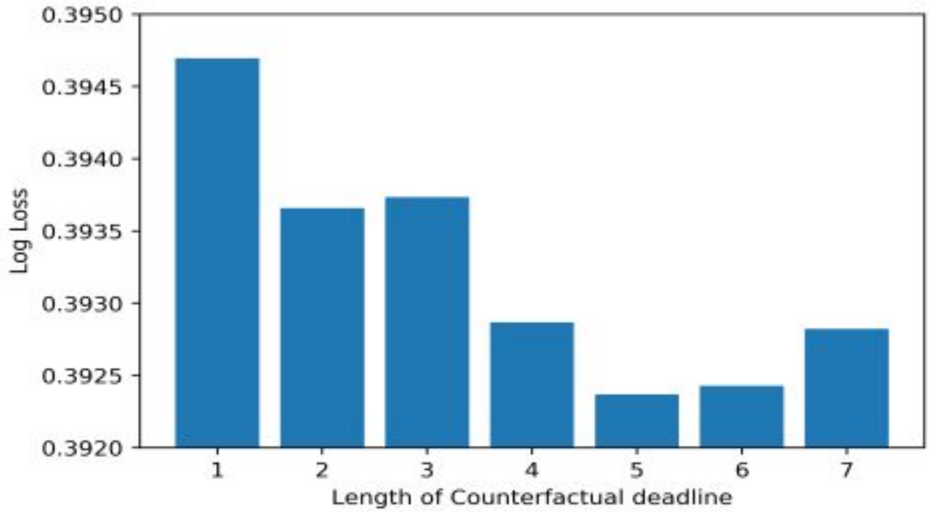


Figure 5: LL of different counterfactual deadline lengths