# Delayed Feedback Model with Negative Binomial Regression for Multiple Conversions

Youngmin Choi, Mugeun Kwon, Younjin Park, Jinsoo Oh, Suyoung Kim

LINE PLUS

LINE

# OUTLINE

INTRODUCTION

PROPOSED MODEL

EXPERIMENT RESULTS

RECAP

# INTRODUCTION

# PROBLEM AND CHALLENGE

## Goal

- Predict the **eventual number of conversions** accurately.
- Input X: feature(User, Site, Ad), Target Y: the number of conversions

## Challenge

- **Long delay** between click and conversions
- **Multiple occurrences** of conversions

| Conversion Type | Non-binary ratio(%) |
|---|---|
| Type A | 26.4% |
| Type B | 1.1% |
| Type C | 0.0% |
| Type D | 59.3% |

## Contribution

- Cover multiple conversions and delay simultaneously
- Reasonable and easy to be implemented

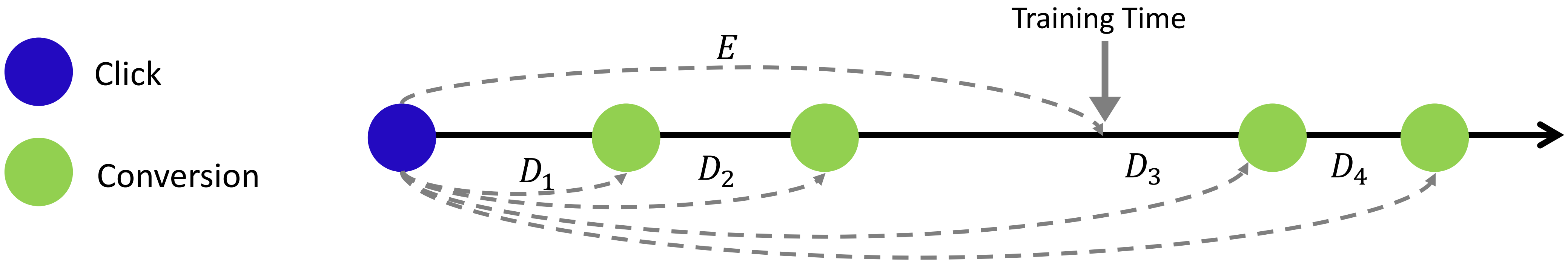
# RELATED WORK

## Modeling Delayed Feedback in Display Advertising (Chapelle, 2014)

| | Conversion | Delay |
|---|---|---|
| DFM | Binary<br>Logistic Regression | An exponential distribution |
| NBDFM<br>(Our Model) | Countable<br>Negative Binomial Regression | Order Statistics |

# PROPOSED MODEL

# PROPOSED MODEL - Notation




$X$: a set of features

$Y \in \{0, 1, \dots\}$ : the number of conversions at the current time

$C \in \{0, 1, \dots\}$ : the number of eventual conversion

$D_k$ : the delay between click and the $k^{th}$ conversion

$E$ : the elapsed time since the click

$Y = 2$

$C = 4$



LINE

# PROPOSED MODEL - NBDFM

## Objective Function

- Negative log-likelihood of observed samples
- Suppose we observe n samples, $\left\{x_j, a_j, d_{j1}, \dots, d_{ja_j}\right\}_{j=1}^{n}$
- $L(w_c, w_d) = -\sum_{j=1}^{n} \log(\Pr(Y_j = a_j, D_{j1} = d_{j1}, \dots, D_{ja_j} = d_{ja_j} | X_j = x_j, E_j = e_j))$

## Optimization

- Estimate parameters by minimizing our objective function
- Use L-BFGS as an optimizer

## Prediction

parameter vector

- $E(C|X = x) = \frac{p(x)}{1-p(x)}$, where $p(x) = (1 + \exp(-\boxed{w_c}x))^{-1}$

# PROPOSED MODEL - Models

## Probability of conversion $\boldsymbol{Pr(C = k \mid X = x)}$

- A standard negative binomial regression model
- $Pr(C = k|X = x) = p(x)^k(1 - p(x))$ with $p(x) = (1 + exp(-\boxed{w_c}x))^{-1}$

parameter vector

## Joint probability of delays between click and multiple conversions $\boldsymbol{D_1, \dots, D_k}$

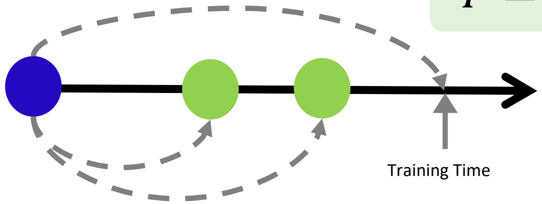
- Order statistics of i.i.d random variable following an Exponential distribution
- $Pr(D_1 = d_1, \dots, D_k = d_k|X = x, C = k) = k!\ \lambda(x) \exp\left(-\lambda(x) \sum_{i=1}^{k} d_i\right)$ with $\lambda(x) = exp(\boxed{w_d}x)$

parameter vector

# PROPOSED MODEL - Sketch to calculate log-likelihood

$$Y = a \iff C = a \text{ or } (C = k \text{ and } E < D_{a+1} < \ldots < D_k \text{ for all } k > a)$$

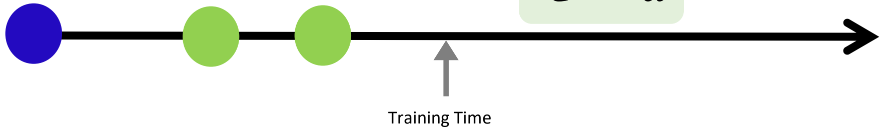
$Y = a$




$X = x$
$E = e$
$D_1 = d_1, \ldots, D_a = d_a$

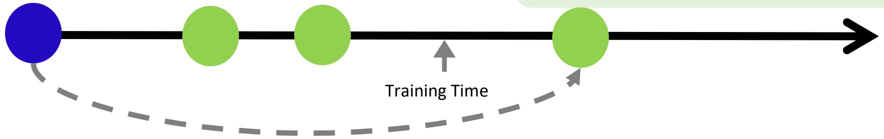
$\Pr(C = a|X = x)$

$C = a$



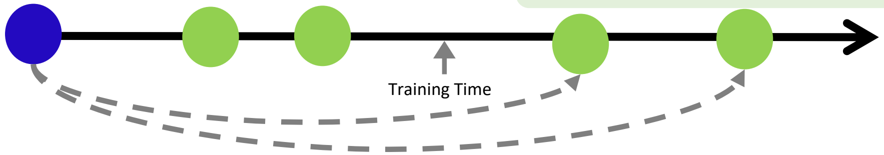

$\Pr(C = a + 1|X = x)$

$C = a+1, E < D_{a+1}$




$\Pr(C = a + 2|X = x)$

$C = a+2, E < D_{a+1} < D_{a+2}$




...

# PROPOSED MODEL - Sketch to calculate log-likelihood

line 1 $Pr(Y = a, D_1 = d_1, \ldots, D_a = d_a | X = x, E = e)$

line 2 $= \sum_{k=a}^{\infty} \boxed{Pr(Y = a, D_1 = d_1, \ldots, D_a = d_a | C = k, X = x, E = e)} Pr(C = k | X = x)$

line 2-a

By $Y = a \iff C = a$ or $(C = k \text{ and } E < D_{a+1} < \ldots < D_k \text{ for all } k > a)$

line 2-a $Pr(Y = a, D_1 = d_1, \ldots, D_a = d_a | C = k, X = x, E = e)$

By $Y = a$ given $C = k$ $\Rightarrow (E < D_{a+1} < \ldots < D_k)$

line 2-b $= Pr(D_1 = d_1, \ldots, D_a = d_a, E < D_{a+1} < \cdots < D_k | C = k, X = x, E = e)$

By the models for delays

line 2-c $= k!\, \lambda(x)^a exp(-\lambda(x) \sum_{i=1}^{a} d_i) \int_e^{\infty} \cdots \int_{d_{k-1}}^{\infty} \lambda(x)\, exp(-\lambda(x) d_k) \cdots d(d_k) d(d_{a+1})$

line 2-d $= a!\, \lambda(x)^a exp(-\lambda(x) \sum_{i=1}^{a} d_i) \binom{k}{k-a} exp(-(k-a)\lambda(x)e)$

line 3 $= \sum_{k=a}^{\infty} a!\, \lambda(x)^a exp(-\lambda(x) \sum_{i=1}^{a} d_i) \binom{k}{k-a} exp(-(k-a)\lambda(x)e) p(x)^k (1 - p(x))$

By the negative binomial theorem

line 4 $= a!\, \lambda(x)^a exp(-\lambda(x) \sum_{i=1}^{a} d_i) (1 - p(x)) \boxed{\sum_{j=0}^{\infty} \binom{j + (a+1) - 1}{j} (p(x)\, exp(-\lambda(x)e))^j} \rightarrow \boxed{(1 - p(x)\, exp(-\lambda(x)e))^{-(a+1)}}$

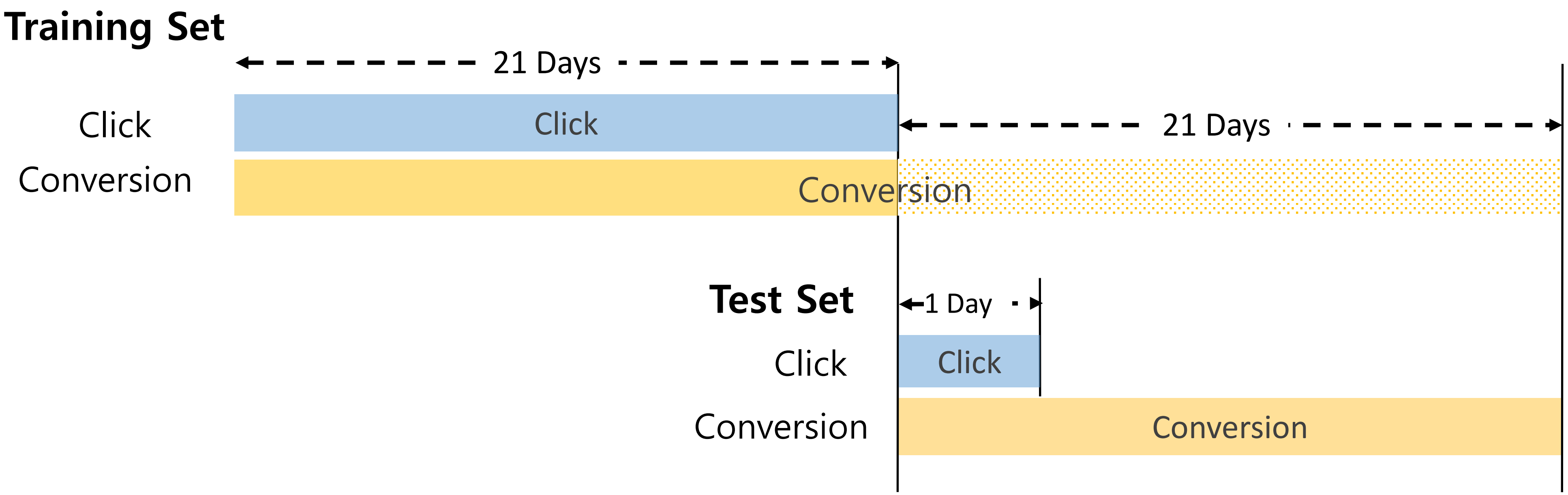


LINE

# EXPERIMENT RESULTS

# EXPERIMENT RESULTS - Dataset Settings

- **No public dataset** for evaluate multiple conversions
- Thus, we used LINE's **real-traffic** logs
- 21 days for training set, following day as test set
- **Repeated 7 times** to check consistency of the results






LINE

# EXPERIMENT RESULTS - Metrics and Competing Models

## Evaluation Metrics

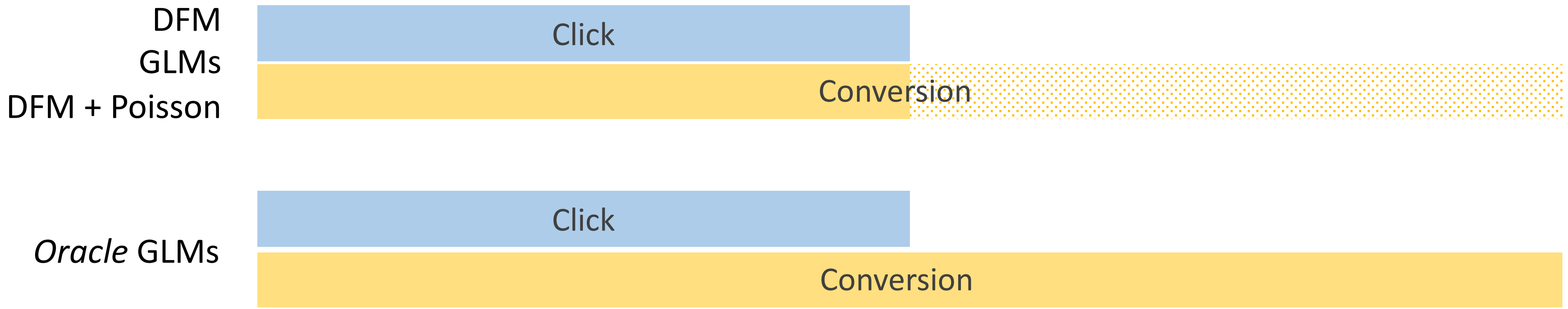
- Mean Squared Error (MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{j=1}^{n}(c_j - E_j)^2$$

- Calibration

$$\mathrm{Calibration} = \frac{\sum c_j}{\sum E_j}$$

## Training Data




## Competing Models

- DFM
- GLMs
  - Logistic / Poisson / Negative Binomial
- DFM + Poisson: Heuristic
- Oracle GLMs: Upper bound

# EXPERIMENT RESULTS - Results

## Overall weighted metric

- Results of different models on 7 test days
- The column 'Diff' shows the difference of MSE between the given model and DFM

| | | MSE | Diff | Calibration(%) |
|---|---|---|---|---|
| Delay only | DFM | 0.09219 | | 141.77 |
| | Logistic | 0.09231 | -0.00012 | 146.60 |
| Count only | Poisson | 0.08681 | 0.00537 | 108.85 |
| | Negative Binomial | 0.08682 | 0.00536 | 108.19 |
| | DFM + Poisson | 0.08723 | 0.00496 | 106.25 |
| Count + Delays | NBDFM | 0.08454 | 0.00764 | 101.12 |
| | *Oracle* Logistic | 0.09223 | | 140.82 |
| Upper bound | *Oracle* Poisson | 0.08298 | | 100.34 |
| | *Oracle* Negative Binomial | 0.08248 | | 99.29 |

# RECAP

Multiple Conversions

Multiple Delays

Together

# RECAP

- Solve the real-world conversion prediction problem, the **key to successful RTB auction**
- Introduce a method **NBDFM** which allows to model both **multiple conversions and delays** jointly
    - Negative Binomial
    - Order Statistics
- Achieve a **significant improvement** in experiments
- Powerful but **simple to deploy** on production services

# Q&A

Thanks