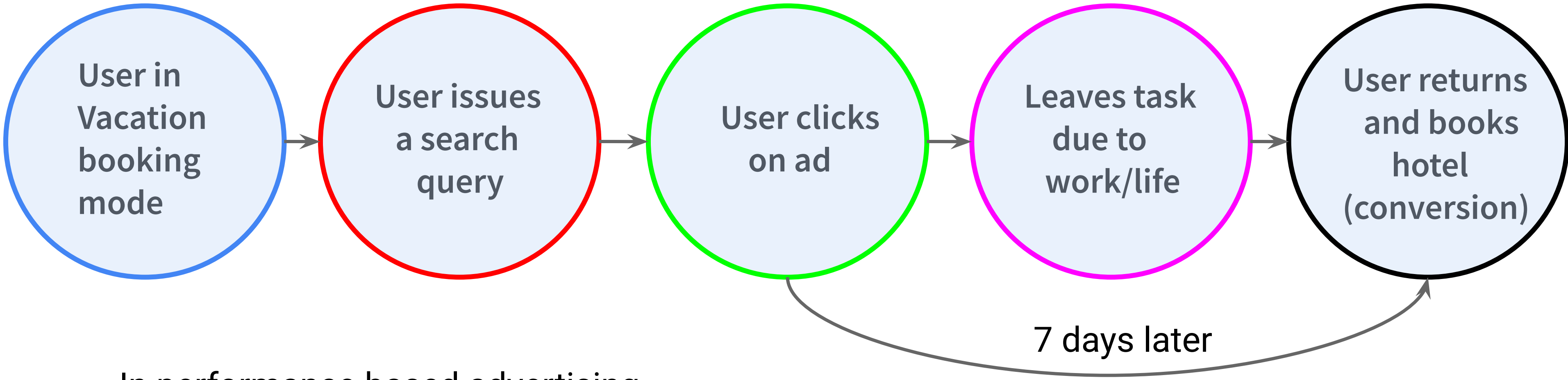
Handling many conversions per click in modeling delayed feedback

Ashwinkumar Badanidiyuru	**Google Research**
Andrew Evdokimov	Google, Inc
Vinodh Krishnan	Google, Inc
Pan Li	Google, Inc
Wynn Vonnegut	Google, Inc
Jayden Wang	Google, Inc

ADKDD 2021

Conversion delay - Problem definition



- In performance based advertising
  bid ~ E[conversions | click] or E[value of conversions | click]
- Supervised machine learning model
  Label = Conversions after the click, attribute conversions to most recent click
- Underprediction due to label from most recent clicks not being mature.

$$\text{Bias} = \frac{\text{predicted conversion rate}}{\text{true conversion rate}} - 1$$

Problem studied - Train a supervised machine learning model while fixing bias due to immature labels

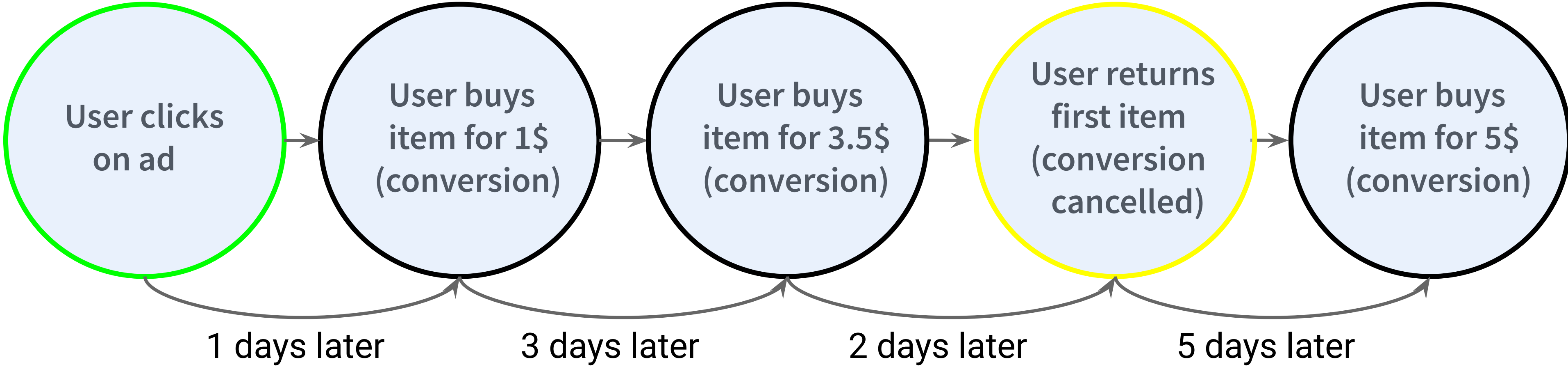
Existing literature

- Problem introduced by a very nice paper by Olivier Chapelle 2014
  Modeling Delayed Feedback in Display Advertising
  Solves the problem via survival analysis

- Follow up literature
  Hubbard et al 2019, Ktena et al 2019, Yoshikawa et al 2018, Mann et al 2019, Wang et al 2017, Safari et al 2017, Vernade et al 2017, Saito et al 2020, Yasui et al 2020, Su et al 2020, Kato et al 2020 etc …

- Most relevant follow up paper
  Delayed Feedback Model with Negative Binomial Regression for Multiple Conversions.
  Choi et al 2020.
  Extends Olivier Chapelle’s results to use survival analysis for multiple conversions per click
  Doesn’t work for our setting.

Differences with respect to previous papers



- Label
  - Multiple conversions per click
  - Delay distribution doesn’t follow a family (is not exponential)
  - Float numbers for value of conversions
  - Retractions of conversions due to user returning an item

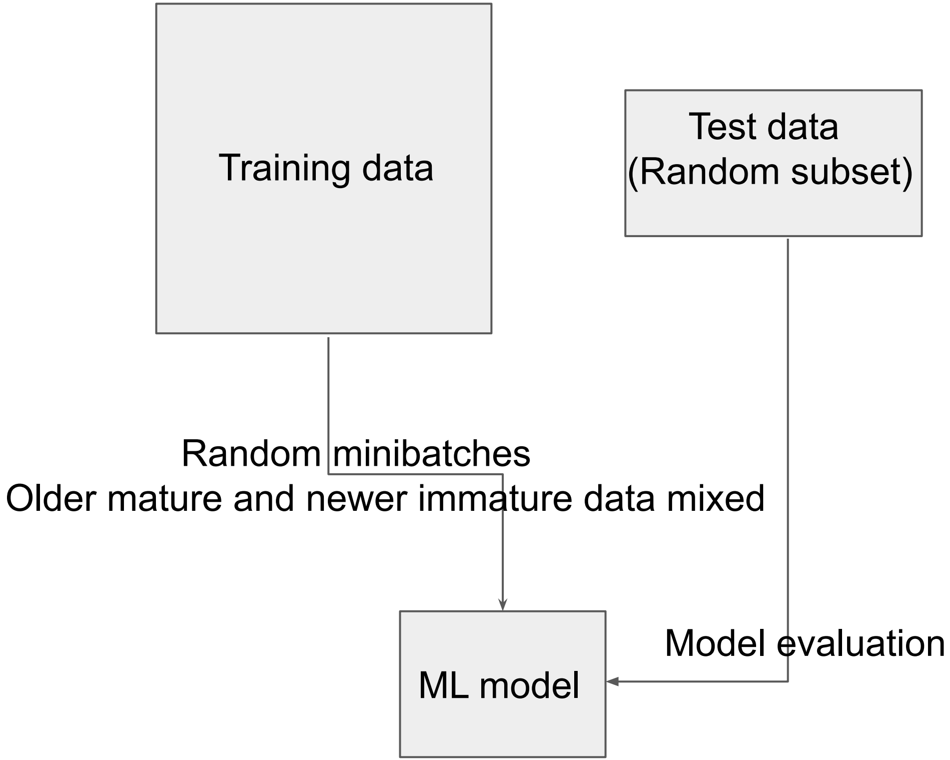
Previous paper

Current paper

Differences with respect to previous papers

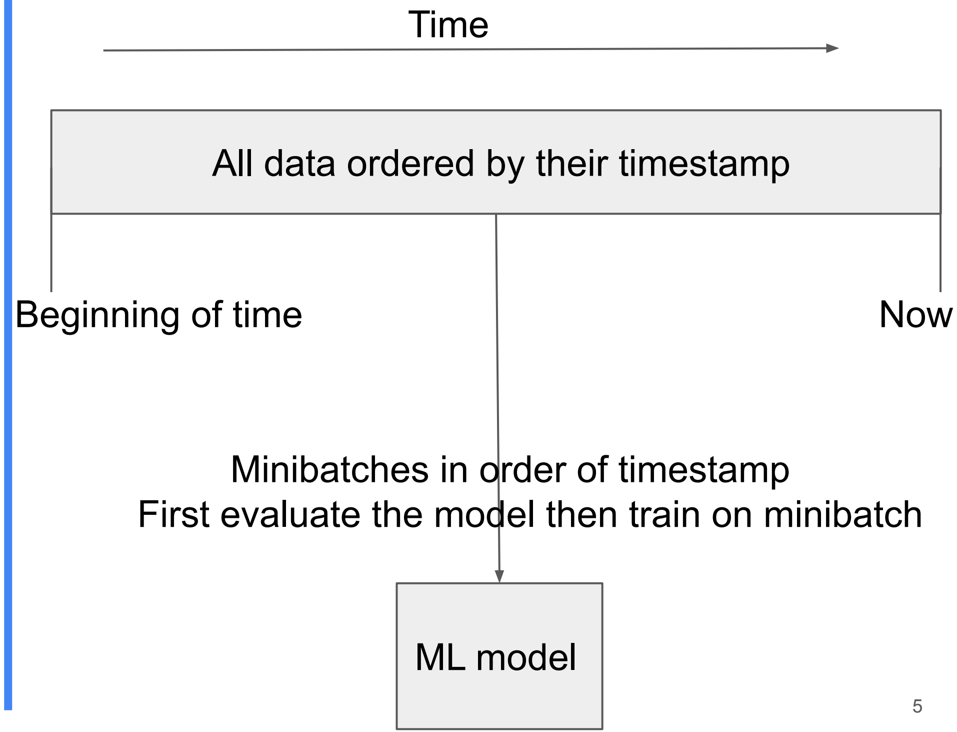
Batch learning

Previous papers



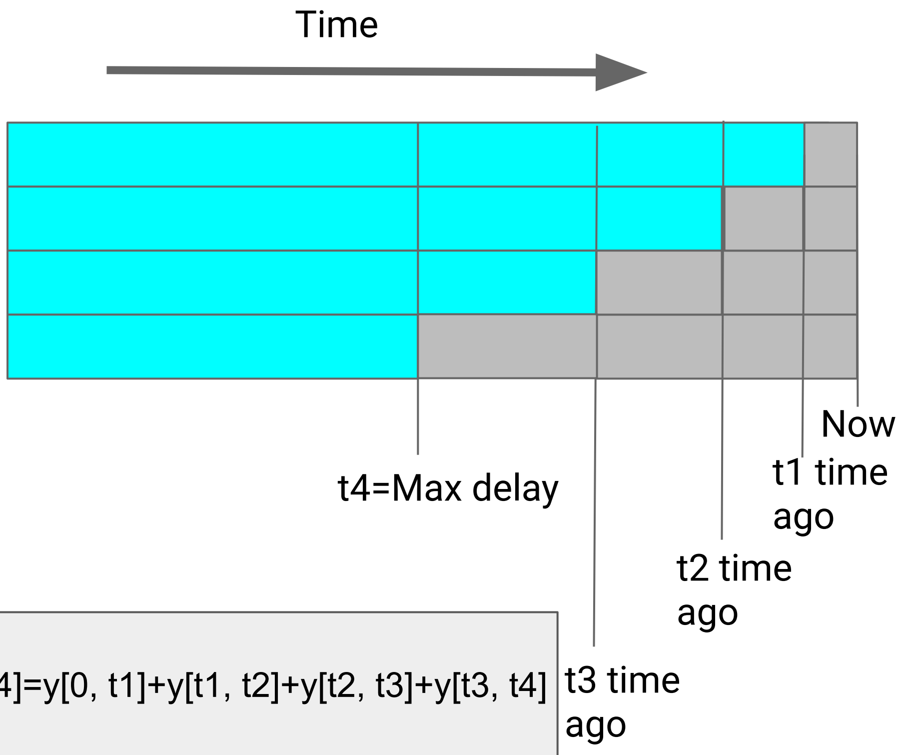
Online learning

Current paper



Proposed model

Idea 1 - Split the label into different delay buckets



Model 4 - Conversions in between day t3 and t4
Trains on examples t4 time ago
Label = y[t3, t4]

Model 3 - Conversions in between day t2 and t4
Trains on examples t3 time ago
Label = y[t2, t3]

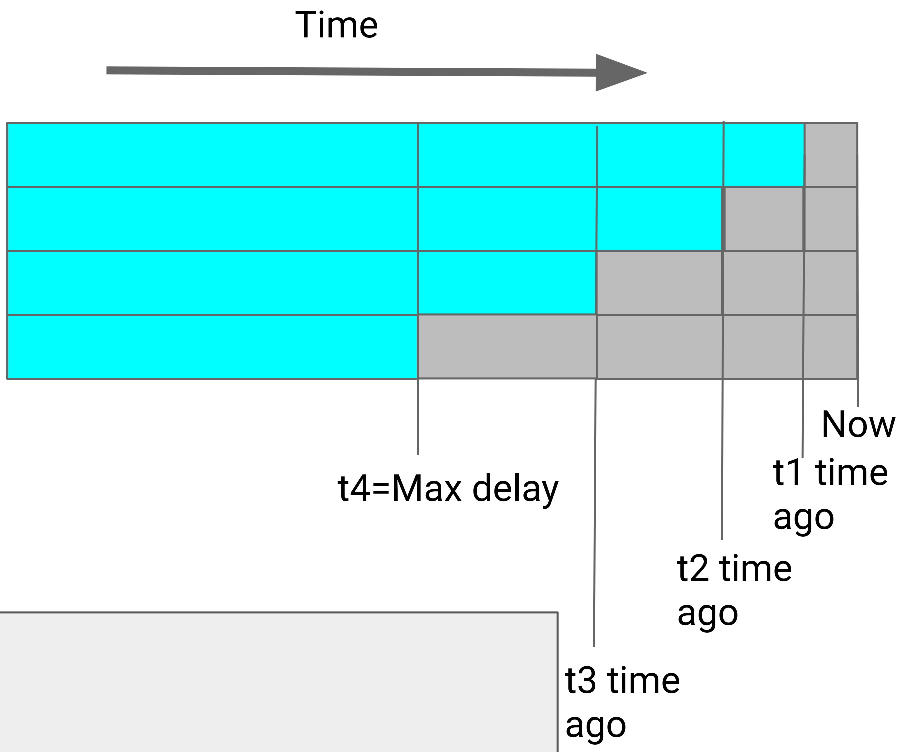
Model 2 - Conversions in between day t1 and t4
Trains on examples t2 time ago
Label = y[t1, t2]

Model 1 - Conversions in between day 0 and t4
Trains on examples t1 time ago
Label = y[0, t1]

y[0, t4]=y[0, t1]+y[t1, t2]+y[t2, t3]+y[t3, t4]

Prediction = Prediction Model 1 +
Prediction Model 2 +
Prediction Model 3 +
Prediction Model 4

Idea 2 - Thermometer encoding



Model 4 - Conversions in between day t3 and t4
Trains on examples t4 time ago
Label = y[t3, t4]

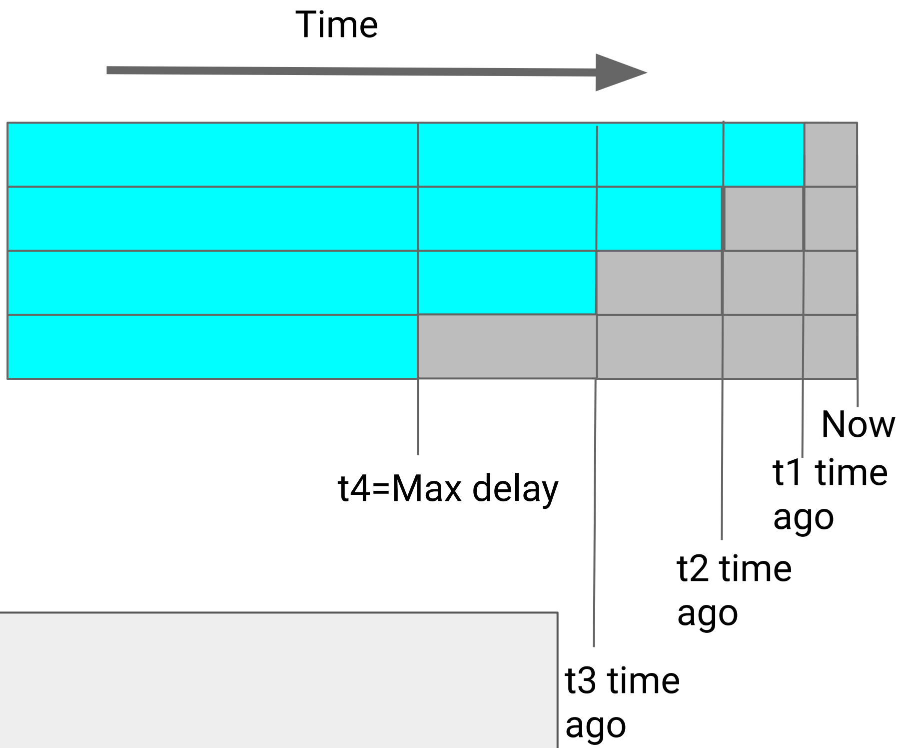
Model 3 - Conversions in between day t2 and t4
Trains on examples t3 time ago
Label = y[t2, t3] + Prediction Model 4

Model 2 - Conversions in between day t1 and t4
Trains on examples t2 time ago
Label = y[t1, t2] + Prediction Model 3

Model 1 - Conversions in between day 0 and t4
Trains on examples t1 time ago
Label = y[0, t1] + Prediction Model 2

Prediction = Prediction Model 1

Idea 3 - Auxiliary information



Prediction = Prediction Model 1

Model 4 - Conversions in between day t3 and t4
Trains on examples t4 time ago
Label = y[t3, t4]
<u>Feature includes y[0, t3]</u>

Model 3 - Conversions in between day t2 and t4
Trains on examples t3 time ago
Label = y[t2, t3] + Prediction Model 4
<u>Feature includes y[0, t2]</u>

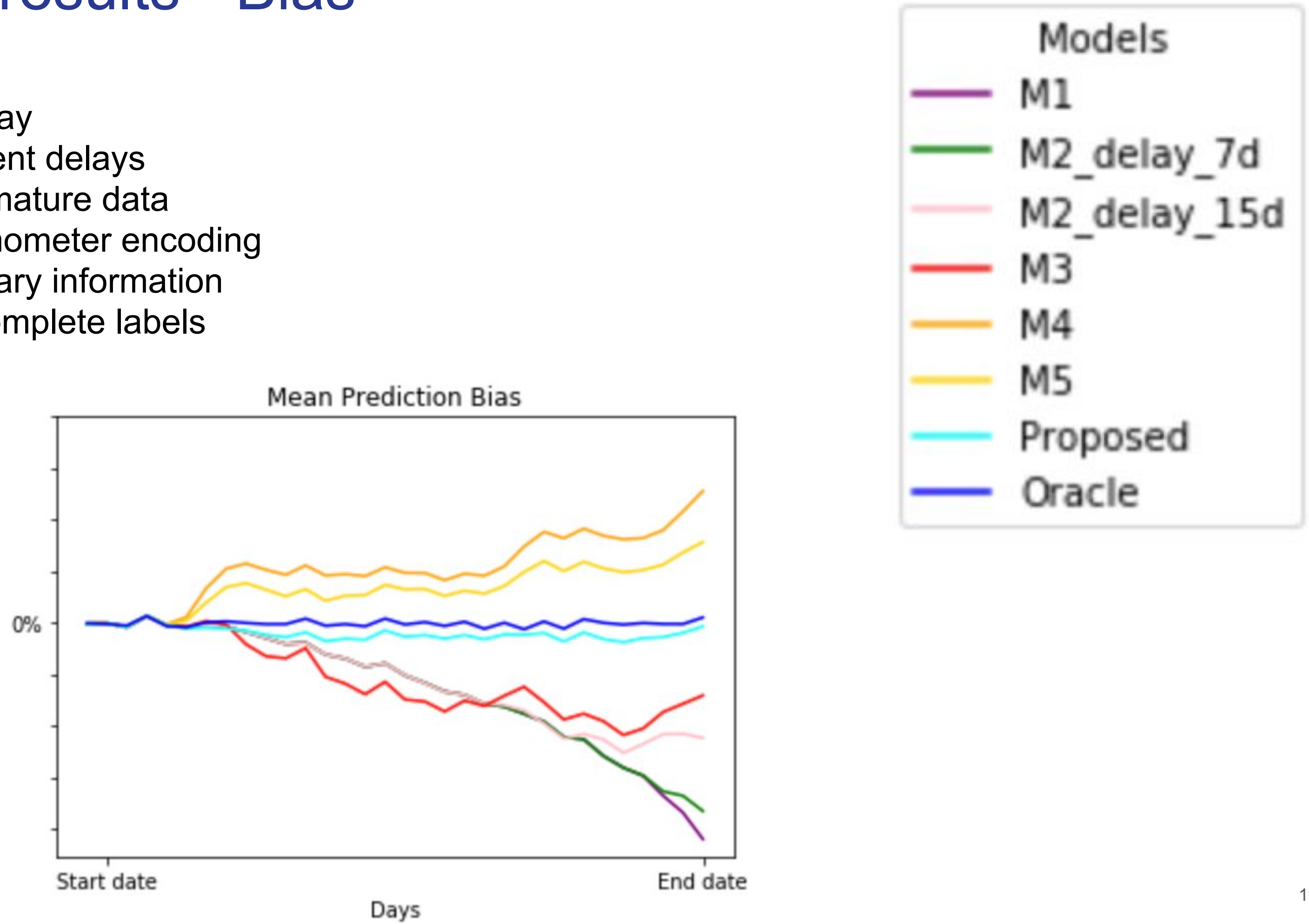
Model 2 - Conversions in between day t1 and t4
Trains on examples t2 time ago
Label = y[t1, t2] + Prediction Model 3
<u>Feature includes y[0, t1]</u>

Model 1 - Conversions in between day 0 and t4
Trains on examples t1 time ago
Label = y[0, t1] + Prediction Model 2

Experimental results

Experimental results - Bias

- M1: Neglecting delay
- M2: Train on different delays
- M3: Train only on mature data
- M4: Remove thermometer encoding
- M5: Remove auxiliary information
- Oracle: Train on complete labels



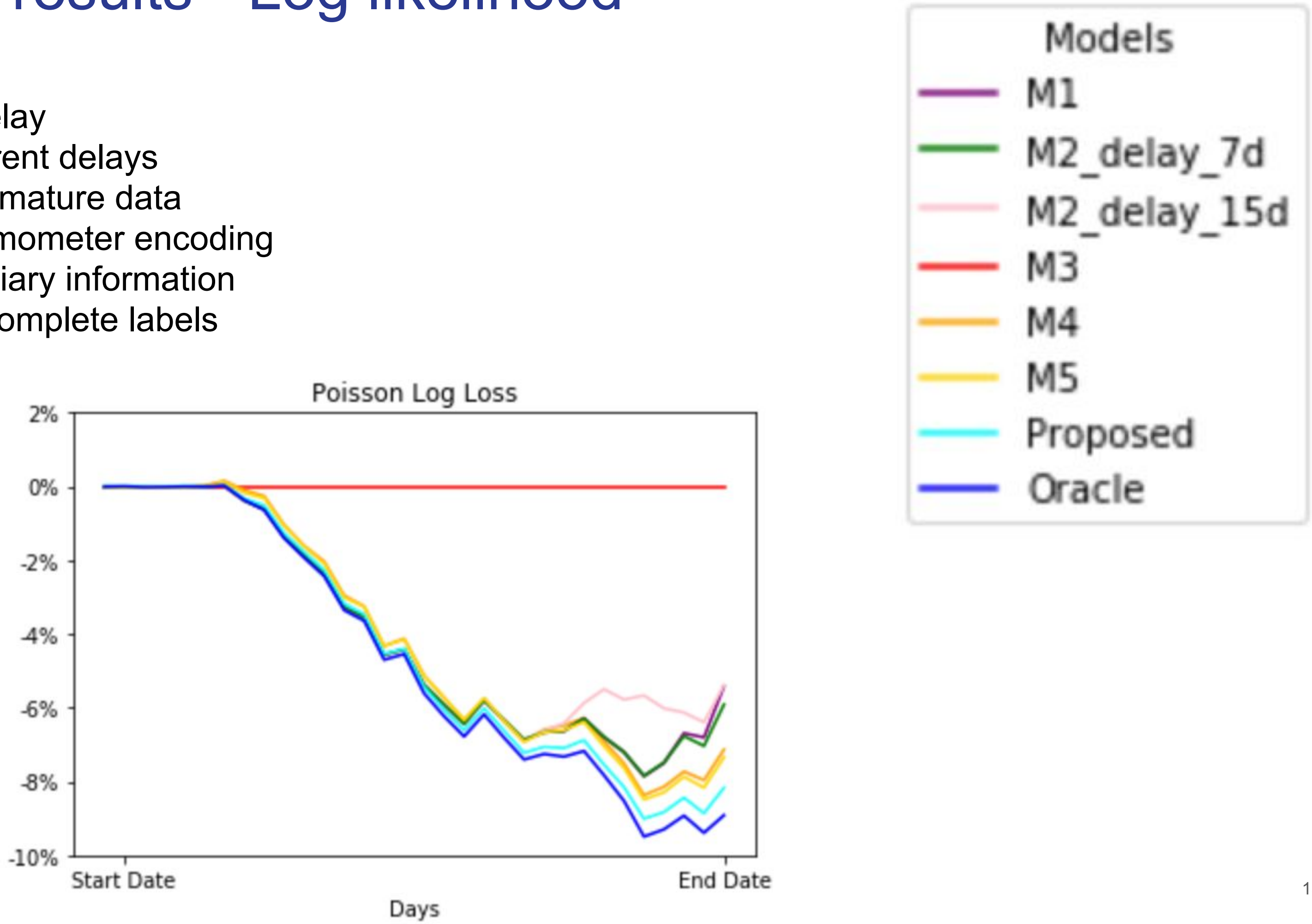
Experimental results - Negative log likelihood

- M1: Neglecting delay
- M2: Train on different delays
- M3: Train only on mature data
- M4: Remove thermometer encoding
- M5: Remove auxiliary information
- Oracle: Train on complete labels

Model	All data	Long delay advertisers	New advertisers
M3	0.0%	0.0%	0.0%
M1	-6.6%	-7.68%	-0.32%
M2_delay_7d	-6.8%	-7.97%	-0.6%
M2_delay_15d	-5.9%	-7.1%	-1.13%
M4	-7.7%	-9.13%	-0.4%
M5	-7.92%	-9.3%	-1.7%
Proposed	**-8.6%**	**-10.16%**	**-1.81%**
Oracle	-9.1%	-10.87%	-2.0%

Experimental results - Log likelihood

- M1: Neglecting delay
- M2: Train on different delays
- M3: Train only on mature data
- M4: Remove thermometer encoding
- M5: Remove auxiliary information
- Oracle: Train on complete labels



Q&A