Attribution Measurement Use Cases

Charlie Harrison
Google
PATCG May 2022

Agenda

- Reporting
- Incrementality
- Optimization
- Third party reporting
- Cross environment attribution
- Possibly for tomorrow: optimization challenges & call to action

Reporting: advertiser insights

*How much "**value**" did my advertisements "**lead to**"?*

Smaller advertisers

- Basic conversion counting
- Conversions are *rare* events

Larger advertisers

- Finer grained drill-downs
- Conversions counts per
  - Campaign
  - Country
  - Device type
  - Conversion type
  - Impression date
  - Publisher site
  - etc.
- Often a huge, sparse key space

Incrementality / Lift

- Run an A/B experiment for a particular campaign advertiser
- Where the advertiser's ad *would have shown up*, divert the user
  - A branch: display the ad normally
  - B branch: show some other ad (or a blank white rectangle)
- Finally: run statistical tests to see if the A branch actually converted more, relative to the baseline B branch
- Causation vs. correlation

Key requirement: measuring "non-existent" ads / ad "opportunities"

Optimization: Setting the scene...

You have $1000

How should you spent it on ads?

Optimization: Goal setting

What are you trying to accomplish with the ad campaign?

- Drive account creation?
- Drive value spent on an ecommerce site?
- Drive users to your online game and play it?

Optimization: Goal setting

Maximize # of new high value users created, where “high value” depends on your goals

- Users that create an account
- #Games played by the user in first X days
- $Value spent by the user in first X days

These goals can be modeled as traditional machine learning tasks!

- Binary classification
- Poisson regression
- Linear regression
- Etc.

Predictions used to adjust ad bids during the auction

Optimization: How is it done today?

- Organize important input (serving-side data) into “features”
  - Contextual
  - Ad specific
  - User
- Label features with the true result / outcome (purchase, sign up, etc)
  - Join the conversion with the features with cookies / ADID
- ML model
  - Generalized Linear Models (such as logistic regression)
  - **More recently, taken over by Deep Neural Networks (DNNs)**
  - Other methods: tree-based (e.g. gradient boosted trees)

Third party reporting

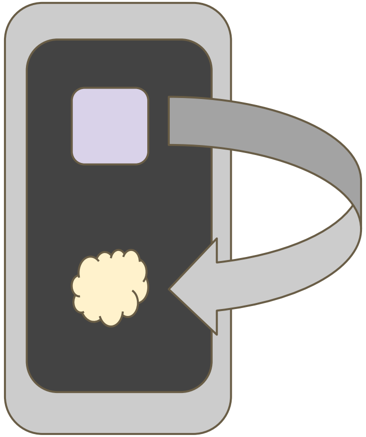
- Advertiser's interact with multiple third parties to perform different tasks
- Examples
    - Ad-tech B is tasked with verifying measurement from Ad-tech A. Both want independent measurement
    - Ad-tech C is tasked with providing more faithful attribution information by observing both Google and Facebook ad-events
    - Ad-tech A and B both specialize in different use-cases, and report the same events differently (incrementality studies vs. basic reporting)

Key problem:

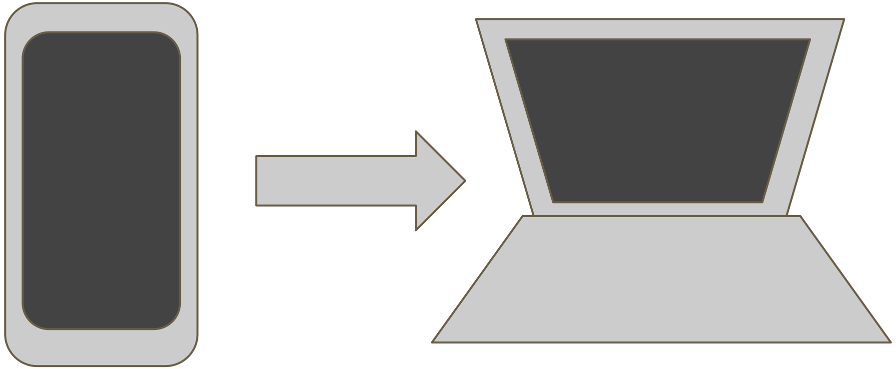
- Multiple third parties need to share a limited resource (privacy budget)
    - large coordination / utility problem
    - Potential denial of service!

Cross environment measurement

App <--> web



Cross device



Optimization: Challenges

Sensitivity management

- How much can a user / event impact the final prediction model / output?
- Different tasks have a different sensitivity, not one-size fits all
- Hugely important consideration in private ML

Noise will scale relative to sensitivity

Sensitivity management: Criteo AdKDD'21 Privacy-Preserving ML Challenge

- https://arxiv.org/pdf/2201.13123.pdf
- eps=5*, delta=e-10
- 19 basic features + 19*18/2 cross features = 190 total features
- Logistic regression implemented via aggregates of the 190 features
- L2 sensitivity bounded by sqrt(190)
  - std=17
- L1 sensitivity (alone) bounded by 190
  - **std=243! Almost 15x more noise!**

Sensitivity management: ML gradient updates

- Some techniques add noise to computed gradients
  - Size of gradients proportional to # params
  - e.g. MaskedLARK, DP-SGD
- Noise scales based on sensitivity of gradients
- Many use cases have a huge number of parameters
  - ~billions
- Without care, noise grows out of control
- Scarcity is common
  - i.e. gradient will be 0 in many places
  - could we take advantage with L0 constraints?

Sensitivity management: conversion values

- Huge variation in max values for advertisers
- How to handle e.g. advertisers that sell a diversity of products
  - Bucketization?
  - Truncation?

10×$(🛹) ± $(🚗) =

🔊

Reporting delay

- Delay to disassociate with sensitive event (attributed conversion)
- Delay due to needing “large enough” aggregates
- Delay due to constrained output domain
    - e.g. sending masked label when an impression “expires”, regardless of whether it actually converted or not

Too much delay can destroy training utility, as recent data is much more informative about predicting upcoming behavior

Negative examples

- In general, supervised learning will need to know negative examples
- Some exceptions
    - e.g. Criteo AdKDD’21 aggregate logistic w/ unlabeled granular data
- Important implications for aggregate data needs!

We need better DP ML research for ads data

Existing research on DP ML differs from our setting in a number of critical ways.

1. Only label privacy needed (in many cases)
   a. In cases where features are "public"
   b. Hugely impactful!
2. Sparse gradient update
3. Prediction types useful for ads
4. Delay

- **It would be great to have more public ads data to benchmark private ML techniques**
- Many questions to answer, even in classical (non-MPC) computational model