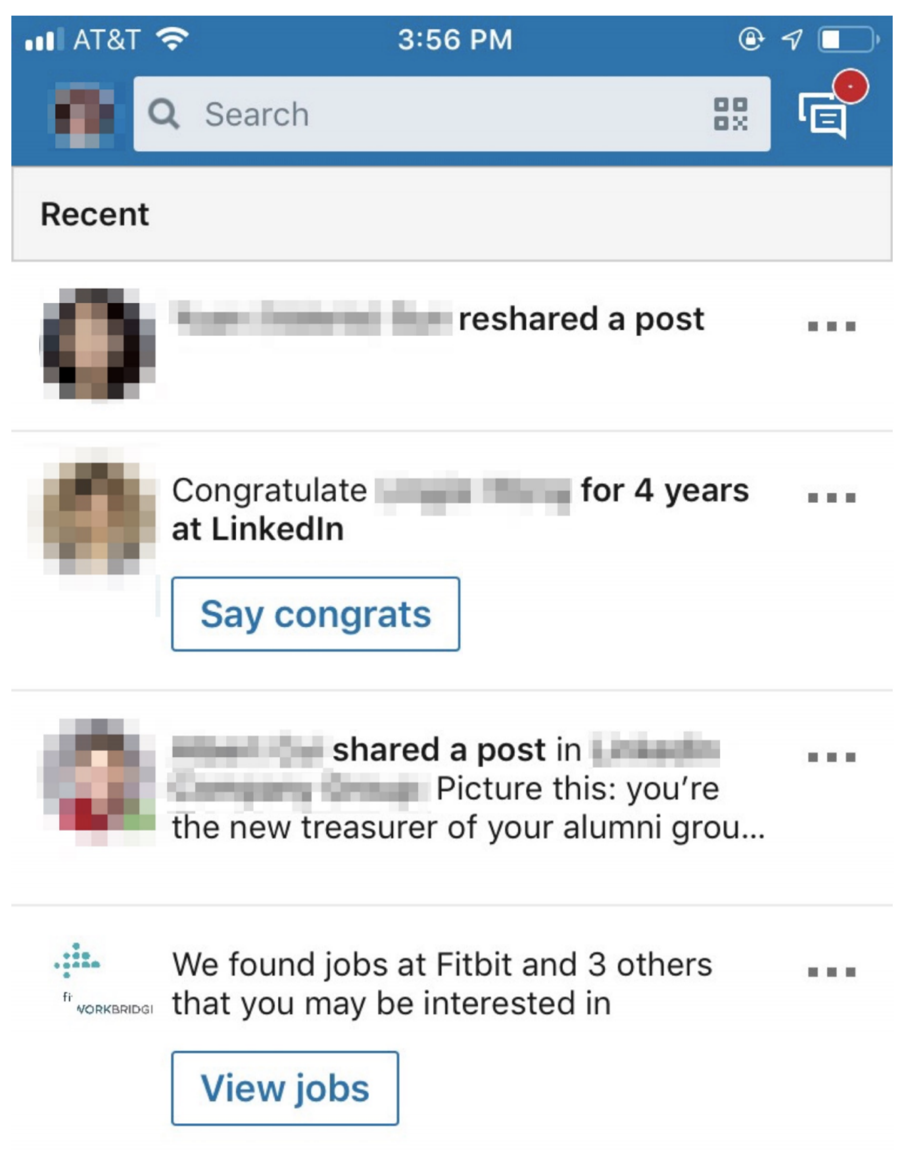
Personalization and Optimization of Decision Parameters via Heterogenous Causal Effects

Ye Tu, Kinjal Basu, Jinyun Yan, Shaunak Chatterjee, Birjodh Tiwana

Introduction and Motivation

In large-scale social media platforms, the member experience is often controlled by certain parameters in conjunction with machine-learned models. Such parameters could be independent of the machine-learned models. An example is the decision to send or drop a notification depending on a threshold parameter, which is used directly on the output of a machine learning model.

If the score of the notification obtained from a machine-learned model is greater than the threshold, the platform then sends the notification otherwise it is dropped (Near real-time optimization of activity-based notifications, Yan et al, 2018). The machine learning model controls the relevance of the item, while the threshold controls certain business metrics.



In an ideal situation, we would often want such parameters to be fully optimized for the multiple business objectives. One option is to fix a global parameter for all members and then iterate on the choice of this parameter through A/B testing. However, that may very well be a suboptimal solution. We propose a new model-based approach that automatically

- Identifies member cohorts with heterogeneous causal effects leveraging a wide range of features.
- Selects the best parameter (e.g. threshold) for each cohort through stochastic optimization to personalize member experience.

Notation and Definitions

Symbol	Meaning
J	Total number of parameter values or choices.
K	Total number of guardrail metrics
C_i	i-th cohort for $i = 1, \ldots, n$.
$\mathbf{U}_k$	Vectorized version of U_{ij}^k, which is the causal effect in metric k by parameter j in cohort C_i.
$\boldsymbol{\mu}_k$	Mean of $\mathbf{U}_k$
$\boldsymbol{\sigma}_k^2\mathbf{I}$	Variance of $\mathbf{U}_k$
$\mathbf{x}$	The assignment vector.

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{Maximize}} \quad & \mathbf{x}^T\mathbf{U}_0 \\
\text{subject to} \quad & \mathbf{x}^T\mathbf{U}_k \leq c_k \qquad \text{for } k = 1, \ldots, K. \\
& \sum_j x_{ij} = 1 \ \ \forall i, \qquad 0 \leq \mathbf{x} \leq 1
\end{aligned}
$$

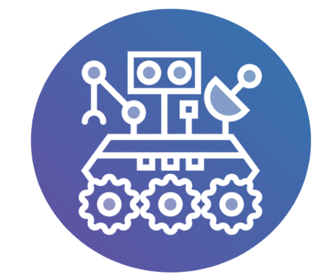
Based on these notations, we can formulate our optimization problem. Let k = 0 denote the main metric. We wish to optimize the main metric keeping the guardrail metrics at a threshold. Formally, we wish to get the optimal x^* by solving the above maximization problem: where c_k are known bounds.

Methodology

Heterogeneous Cohort Identification

Regression Tree (CART)

- Predict Y (sessions)
- Splitting Objective: MSE(Y)



Causal Tree

- Estimate $\boldsymbol{\tau}$ (Δsessions)
- Splitting Objective: MSE($\boldsymbol{\tau}$) + Variance regularizer
- Honest estimations

- Causal tree was introduced by Susan Athey and Guido Imbens's paper "Recursive partitioning for heterogeneous causal effects " in 2013. It is based on the conventional CART algorithm.
- Modifies the splitting tree and cross validation objectives to optimize for the accuracy of estimated delta effects while penalizing the variance in estimations.

Stochastic Multi-objective Optimization

- We start by reframing the optimization problem to a generic form:

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{Minimize}} \quad & f(\mathbf{x}) := \mathbb{E}_{\mathbf{U}_0}\left(F(\mathbf{x}, \mathbf{U}_0)\right) \\
\text{subject to} \quad & g_k(\mathbf{x}) := \mathbb{E}_{\mathbf{U}_k}\left(G_j(\mathbf{x}, \mathbf{U}_k)\right) \leq 0 \qquad \text{for } k = 1, \ldots, K. \\
& x \in \mathcal{X}
\end{aligned}
$$

- Our algorithm, which we call Multiple Coordinated Stochastic Approximation (MCSA), is an iterative algorithm which runs for N steps. At each step t, we start by estimating the constraint function.Specifically, we simulate $U_{k\ell}$ for $\ell = 1, \ldots, L$ and estimate:

$$\hat{G}_{k,t} = \frac{1}{L}\sum_{\ell=1}^{L} G_k(\mathbf{x}_t, \mathbf{U}_{k\ell}).$$

If all constraints satisfied, we optimize towards the objective. Otherwise we randomly pick a constraint, and optimize towards it, and update x.

Algorithm 2 Multiple Cooperative Stochastic Approximation

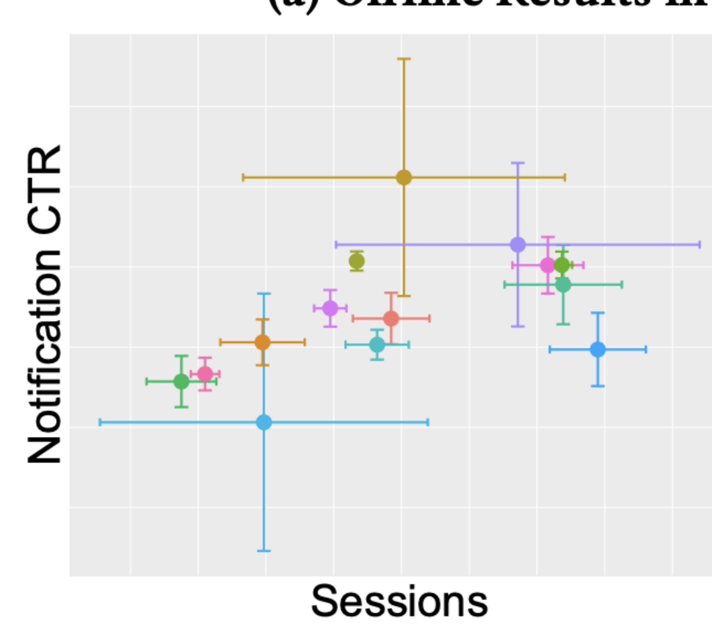
```
Input : Initial x_1 ∈ X, Tolerances {η_k}_t, {γ}_t, Total iterations N
for t = 1, ..., N do
    Estimate Ĝ_{k,t} for all k ∈ 1, ..., K using (6).
    if Ĝ_{j,t} ≤ η_{j,t} for all j then
        Set h_t = F'(x_t, U_{0,t})
    else
        Randomly select k* from {k : Ĝ_{k,t} > η_{k,t}}
        Set h_t = G'_{k*}(x_t, U_{k*,t})
    end if
    Compute x_{t+1} = P_{x_t}(γ_t h_t)
end for
Define B = {1 ≤ t ≤ N : Ĝ_{k,t} ≤ η_{k,t} ∀k ∈ {1, ..., K}}
return x̂ := (Σ_{t∈B} x_t γ_t) / (Σ_{t∈B} γ_t)
```

Offline/Online Evaluations

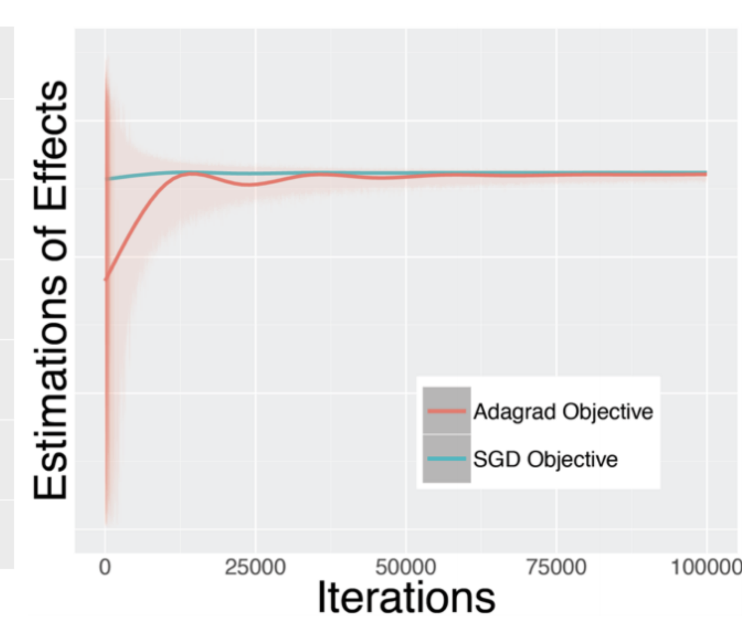
Offline Evaluations. Below figures show the offline estimations.

Online A/B Testings. We demonstrated online validation of the approach which has been tested online on the LinkedIn notification system to increase session/visits metric while holding all guardrail metrics neutral.

(a) Offline Results in LinkedIn Notifications



(a-1) Heterogeneous Cohorts



(a-2) Expectated Sessions Lift

Metrics	Delta % Effects	
	All Users	DAUs
Sessions	+0.37%	+0.56%
Daily Contributors	Neutral	+0.42%
Notification Send Volume	Neutral	Neutral
Notification CTR	Neutral	Neutral
Notification Disable Users	Neutral	Neutral
Notification Total Disables	Neutral	Neutral