Personalized Ad Quality Bidding with MTML Causal Modeling and Constrained Optimization

Libin Liu , **Fang Liu** , **Chen Fu** , **Fei Peng** , **Ethan Shao** , **Wei Guan** , **Kevin Chang**
Meta Platforms, Inc.
{libinliu, fangliu, chenfu, ainikapeng, ethanshao, wguan, kevincha}@meta.com,

Abstract

Ad quality plays a central role in ranking systems, promoting high-quality ads and demoting low-quality ones to enhance user experience and ultimately drive long-term value for both people and businesses. The quality of each ad is estimated by a value model, which computes a weighted sum of various quality predictions. Since different user cohorts exhibit heterogeneous sensitivities to the same ad, personalization aims to customize these weights to achieve a more efficient trade-off between ads performance and user engagement. In this paper, we propose a new personalization framework with two key innovations: 1) a multi-task multi-label (MTML) causal model that jointly predicts user sensitivities across multiple ad quality signals; and 2) a user sensitivity information aware and structural information aware optimization framework for learning more efficient scalar weights. With these improvements, our framework achieves a 0.5% increase in ads performance while maintaining neutral engagement, and delivers a 1.4x gain in efficiency compared to the current system.

1 Introduction

Social media platforms prioritize delivering high-quality user experiences by facilitating the creation and discovery of organic content. To sustain their business models, these platforms strategically insert advertisements into the content stream. While increasing the number of ad impressions may improve short-term ads performance metrics, it can degrade long-term user satisfaction—particularly when users are shown low-quality or irrelevant ads. Negative reactions such as reporting, hiding, or disliking ads are clear signals of dissatisfaction and offer valuable supervision signals to guide ranking decisions. These signals motivate the development of quantitative methods to assess and enforce ad quality as a central part of ad delivery systems.

Modern ad ranking systems use machine learning models to estimate the likelihood of various negative user interactions, such as low relevance scores, high hide rates, or poor feedback. Each model prediction is associated with a positive scalar weight that reflects the system's prioritization of different types of negative signals. The weighted predictions, often referred to as quality bids, are aggregated by a value model to produce the final ad score (Figure 1). This score is used to rank and serve ads, balancing ads performance goals with user experience constraints. The effectiveness of this balance directly affects both platform ads performance and user retention, making the value model a central component of any large-scale ad system.

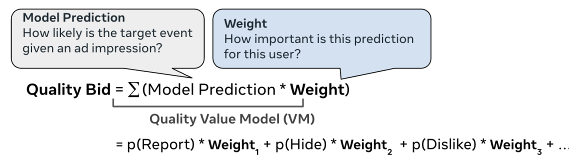


Figure 1: Illustration of quality value model.

Traditionally, the value model applies a uniform policy in which the scalar weights are fixed across the entire user base. However, users exhibit substantial heterogeneity in their sensitivities to different ad signals. For instance, some users may be highly tolerant of repetitive creatives or strong calls-to-action, while others may find such features intrusive. Applying the same penalty weights to all users fails to capture these nuances, potentially leaving performance gains on the table. By customizing scalar weights for different user cohorts, the system can selectively relax or enforce quality constraints, thus enabling more personalized ad exposure strategies. We refer to this scalar customization process as *personalization*[1].

The personalization process consists of two core components: cohort prediction and scalar weight optimization. In the first stage, the system must infer which cohort a given user belongs to based on their sensitivity to various ad quality signals. Cohort prediction is a crucial step tailoring the ad experience to users in different cohorts, it can be achieved through various machine learning models, such as clustering algorithms (*e.g.* k-means [MacQueen and B., 1967]) or classification models (*e.g.* logistic regression [Jr. *et al.*, 2013]), these methods typically group users by their demographic, behavior and engagement pattern. However, these classical machine learning models may not capture the underly-

[1] Ideally we want the personalization process to be performed on individual user level, due to complexity constraints, currently we focus on user cohort level.

ing causal relationships between user characteristics and ad preferences. To address this limitation, causal modeling techniques are employed, allowing us to identify the causal relationships between variables rather than correlations and enabling more accurate predictions and better decision-making. The key focus of causal modeling for cohort prediction is the incremental change in outcome due to a treatment or intervention [Rubin, 2005], which is termed heterogeneous treatment effect (HTE) or conditional average treatment effect (CATE). Uplift models have been used extensively in the analysis of HTE/CATE and there are several major categories: (1) Causal Trees, where different splitting criteria such as distribution divergences [Radcliffe and Surry, 2011] and expected responses [Saito *et al.*, 2020; Zhao *et al.*, 2017] are used to divide user groups; (2) Meta-learner, which predicts the expected outcome to tackle the counterfactual problem [Künzel *et al.*, 2019; Alaa *et al.*, 2023]; (3) Deep Neural Networks (DNNs), which leverage the robust representation power and the exceptional predictive capability of DNNs to model causal relationships [Shi *et al.*, 2024; Raul *et al.*, 2023; Shalit *et al.*, 2017; Louizos *et al.*, 2017]; and (4) Sequential modeling, which focuses on long-term reward and enables continuous learning [Zhao *et al.*, 2020; Du *et al.*, 2019a]. In our proposed framework, we design a multi-task multi-label (MTML) causal model that jointly predicts user sensitivity across multiple quality dimensions. This model captures shared representations across tasks while modeling the individual treatment effects for each quality signal. As a result, it provides a compact and expressive way to encode user-level quality preferences.

The second component of our personalization framework involves optimizing the scalar weights for each user cohort. This is a constrained optimization problem that seeks to maximize user engagement subject to ad performance constraints. There are various optimization problem formulation to achieve this goal, such as Bayesian optimization [Agarwal *et al.*, 2018], policy gradient approach [Jeunen *et al.*, 2024], multi-objective optimization [Tang *et al.*, 2024] using convex optimization [Boyd and Vandenberghe, 2004], learning based control [Agarwal *et al.*, 2014], functional optimization [Zhang *et al.*, 2014] and reinforcement learning [Cai *et al.*, 2017]. However, many of these approaches—especially black-box or sample-inefficient methods such as Bayesian optimization and reinforcement learning—face significant practical challenges when applied at industrial scale. These include high computational cost, lack of transparency, lower robustness and slow convergence. In contrast, we adopt a white-box convex optimization approach, which provides a more tractable and interpretable solution pathway. Convex programs enable deterministic guarantees, exploit the problem structure induced by the value model and user sensitivity estimates, and scale efficiently to the billions of impressions handled in a production system. This formulation ensures both robustness and deployability, making it well-suited for large-scale personalization.

In this paper, we present a novel bi-level personalization framework called Opus (OPtimization and User Sensitivity modeling); the detail flow of this framework are: 1) in the user cohort prediction phase, a MTML model is developed to predict user's sensitivity level for multiple quality bids simultaneously; 2) in the mathematical optimization phase: a convex optimization problem is formulated and solved to obtain the optimal scalar for each user cohort. The contributions of this paper are as follows:

- We design a multi-task multi-label (MTML) causal model that jointly estimates user sensitivity across multiple ad quality dimensions. By leveraging cross-domain knowledge transfer, the model effectively addresses quality signal sparsity and achieves an average AUCC gain of 2% (Area Under Cost Curve).
- We propose a convex optimization framework that learns personalized value model weights, enabling fine-grained control over the trade-off between user engagement and ad performance. This approach improves engagement/ads performance metric efficiency by 1.44× compared to the current baseline.
- Our end-to-end system delivers a 0.5% lift in ad performance while maintaining neutral user engagement, demonstrating measurable impact on one of the world's largest social media platforms.

2 Preliminaries and Related Work

We begin with a high-level overview of the Opus framework; detailed architectural and modeling components are presented in the following section.

2.1 Heterogeneous Treatment Effect Modeling

Heterogeneous treatment effect (HTE) modeling refers to the process of predicting a user's sensitivity to changes in ad quality. Typically, for each quality bid, a separate machine learning model—such as a decision tree or deep neural network—is trained to map user features to a cohort identifier. However, this approach has several limitations: 1) The number of models grows linearly with the number of quality bids, making system maintenance increasingly inefficient; 2) Quality bids often exhibit semantic or behavioral similarity, yet independent modeling fails to leverage shared information across tasks, potentially missing out on richer representations and improved performance.

To address these limitations, we adopt a multi-task multi-label (MTML) model paradigm that predicts user cohorts across multiple quality bids simultaneously. MTML architectures have demonstrated strong performance in various ad-related tasks, including personalized recommendation [Gao *et al.*, 2024; Tang *et al.*, 2020; Ma *et al.*, 2018a], model debiasing [Ma *et al.*, 2018b; Zhang *et al.*, 2020], and auction design [Ma *et al.*, 2022; Kalra *et al.*, 2023].

In our MTML setup, all user features are first processed by shared layers to learn generalizable representations, which are then passed to task-specific heads that independently predict cohort assignments for each quality bid. This joint training strategy improves model robustness and overall prediction performance across domains, helping to mitigate common challenges such as data imbalance, noisy supervision, and domain-specific overfitting.

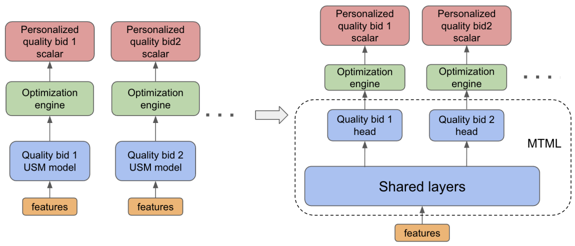


Figure 2: MTML model for user sensitivity prediction.

2.2 Mathematical Optimization

Given the user sensitivity predictions from the MTML model, the next step is to derive optimal scalar weights for each cohort to maximize top-line metrics. To enable supervised learning in this phase, we conduct randomized controlled trial (RCT) experiments. In these experiments, users are randomly assigned to treatment groups, each configured with a different manually selected scalar for a given quality bid. By recording the resulting ad performance and user engagement metrics under each scalar, we construct a training and evaluation dataset for optimization.

Using this data, we formulate a constrained optimization problem as follows:

$$\begin{aligned} \text{maximize} \quad & \text{Total Engagement} \\ \text{subject to} \quad & \text{Total Ads Performance} \geq \text{Performance Constraint.} \end{aligned} \tag{1}$$

In this high-level formulation, the optimization variables are the scalar weights assigned to each user cohort. The objective is to maximize overall user engagement, subject to maintaining a minimum level of ad performance. While alternative formulations are possible (e.g., maximizing ad performance subject to engagement constraints), we adopt the engagement-maximization form in Eq. (1), and demonstrate that it can also be adapted to prioritize ad performance as needed.

3 Detailed Design

In this section, a detailed walk through of the Opus framework will be provided. For notational clarity, matrices are in bold upper case (*e.g.* $\mathbf{A}$), vectors are in bold lower case (*e.g.* $\mathbf{x}$), scalars and variables are non-bold (*e.g.* α, a).

3.1 MTML Model for User Sensitivity Prediction

To estimate the causal treatment effects of multiple quality bids simultaneously, we adopted a unified multi-task multi-label (MTML) causal modeling approach (as shown in Figure 2). Unlike modeling each treatment and domain independently, the MTML framework allows the model to learn shared representations across tasks and domains, facilitating knowledge transfer. This is particularly valuable when data quality or treatment labeling may vary across surfaces—errors or sparsity in one domain can be mitigated by more robust signal from the other. Furthermore, user or contextual patterns may exhibit cross-domain commonalities, which can be effectively captured by shared layers in the MTML architecture, enhancing generalization. By jointly modeling the multiple treatments, we also enable the model to better understand interaction or substitution effects when both treatment types may co-occur or influence similar downstream metrics.

In the MTML architecture, input feature will go through a shared bottom layer, which is responsible for learning common knowledge among all tasks; later the shared output are fed into different task-specific heads which learn task-specific features and produce task outputs. The shared layers and task-specific heads are both feed-forward neural networks with non-linear activation functions.

The input features to the MTML model typically include user demographic features and user activity features, let $\mathbf{x}$ represent the input feature vector and there are N shared layers; also let $\mathbf{W}_i$ and $\mathbf{b}_i$ be the shared weight matrix and shared bias vector in shared layer i, the final output of the shared layer $\mathbf{h}_{shared}$ is represented as:

$$\mathbf{h}_{shared} = \sigma(\mathbf{W}_{N-1}...\sigma(\mathbf{W}_2\sigma\left(\mathbf{W}_0\mathbf{x} + \mathbf{b}_0\right) + \mathbf{b}_1)... + \mathbf{b}_{N-1}), \tag{2}$$

where $\sigma(\cdot)$ is the ReLU activation function. Then the shared layer output $\mathbf{h}_{shared}$ serves as the common input to each task-specific head, similarly, the output of the task-specific layer for task i can be represented as:

$$\mathbf{g}^i = \sigma\left(\mathbf{W}^i\mathbf{h}_{shared} + \mathbf{b}^i\right), \tag{3}$$

where $\mathbf{W}^i$ and $\mathbf{b}^i$ are the weight matrix and the bias vector for task i. After the task layers, the final out is computed as:

$$\mathbf{y}^i = \mathbf{W}^i\mathbf{g}^i + \mathbf{b}^i. \tag{4}$$

To train the MTML model, the loss function is defined as a weighted sum of the mean squared error loss for each task:

$$\mathcal{L}_{total} = \sum_i \omega_i \mathcal{L}(\mathbf{y}^i, \hat{\mathbf{y}}^\mathbf{i}), \tag{5}$$

where

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = -\sum_j (\mathbf{y}_j - \hat{\mathbf{y}}_j)^2. \tag{6}$$

$\hat{\mathbf{y}}$ is ground-truth label vector for each data point and $\mathbf{y}$ is the model output.

In our case, the user's sensitivity score is defined as the expected engagement gain per expected ads performance loss, i.e., the ratio form of two objective CATE's. The larger the ratio, the user is more sensitive to ads quality treatment. However, it is well known that ratio form is more noisy to model directly due to the noise amplification from outliers. To address this challenge, we adopt the Lagrangian selecting criteria[Du *et al.*, 2019b] used in [Shi *et al.*, 2024], to transform the conditional outcome metric to linear form as follows:

$$T_t = R_t + \lambda \cdot Q_t, \tag{7}$$

where R_t and Q_t are the ads performance and quality bid value after treatment t is applied (*e.g.* setting the weighting scalar for this quality bid to a particular value). Note that λ is a hyper-parameter of the model that we will search during the model tuning. Let $t = 1$ represent that certain scalar is applied to the quality bid and $t = 0$ represent the scalar associated to the quality bid is 0, the user's sensitivity (transformed by Lagrangian selecting criteria) can be mathematically computed as:

$$\Delta T = T_{t=1} - T_{t=0}. \tag{8}$$

However, it is challenging to obtain the sensitivity score from the RCT experiment due to counterfactual effect, *i.e.* the individual in the experiment can not be in the treatment and control group simultaneously. To tackle this problem, the doubly robust learner (DRL) method [Shi *et al.*, 2024] is employed to estimate the user sensitivity score; the MTML paradigm is then applied on top of DRL estimator to estimate multiple cross-domain sensitivity score as multiple tasks.

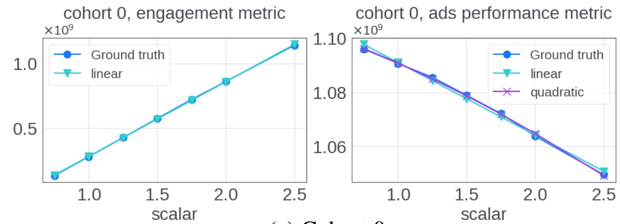
3.2 Linear Programming

Let N be the number of user cohorts and D be the cardinality of scalar decisions, also denote $\{\alpha_0, \alpha_1, ..., \alpha_{D-1}\}$ as the finite scalar decision set that we choose to apply to different user cohorts. From the dataset collected from the RCT experiment; an engagement matrix $\mathbf{E}_{N\times D}$ and an ads performance matrix $\mathbf{P}_{N\times D}$ can be obtained, where $\mathbf{E}_{i,j}$ represents the engagement response of cohort i when applying scalar decision α_j, $\mathbf{P}_{i,j}$ represents the ads performance response of cohort i when applying scalar decision α_j. Let $x_{i,j} \in \{0, 1\}$ be the binary variable denoting whether cohort i should be assigned scalar α_j; also denote B as the baseline ads performance and $0 < \rho < 1$ as the ads performance budget. Given the finite response of scalar decisions, a naive integer programming optimization (ILP) can be formulated as follows:

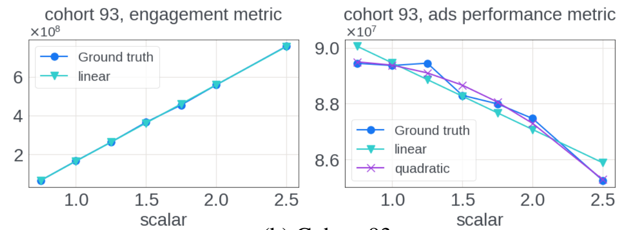
$$\begin{aligned}
\max \sum_{\substack{0\le i<N-1\\0\le j<D-1}} & \mathbf{E}_{i,j} \cdot x_{i,j} \\
\text{s.t} \sum_{\substack{0\le i<N-1\\0\le j<D-1}} & \mathbf{P}_{i,j} \cdot x_{i,j} \ \geq \ (1-\rho)B, \\
\sum_{0\le j<D-1} & x_{i,j} = 1, \ \forall i = 0, 1, ..., N-1.
\end{aligned} \tag{9}$$

The equality constraint in (9) ensures that for each cohort, only one valid scalar is chosen. On one hand, the ILP formulation ensures that the solution is optimal; on the other hand, it is also considered a NP-hard problem, which imposes a great challenge to the scalability as N and D increase; especially when it is more desirable to optimize over a continuous scalar range rather than a finite set of scalar numbers. In order to achieve this goal, certain relaxation is needed to reduce the problem's complexity and some approximations are required to estimate the ads performance and engagement response for each cohort under arbitrary scalar.

In the engagement matrix $\mathbf{E}_{N\times D}$ and ads performance matrix $\mathbf{P}_{N\times D}$, each row of $\mathbf{E}_{N\times D}$ and $\mathbf{P}_{N\times D}$ can be viewed as the engagement response sample points and ads performance response sample points given scalar samples $\{\alpha_0, \alpha_1, ..., \alpha_{D-1}\}$, respectively. With these information, least square curve fitting can be performed to predict the ads performance and engagement behavior of each cohort in the continuous space. A simple approach is to assume linearity in both ads performance and engagement and perform linear fitting. In this scenario, the engagement response for cohort i can be represented as $k_i x_i + d_i$, where k_i and d_i are the slope and intercept of the engagement response for cohort i, x_i is the continuous scalar associated with cohort i; similarly, the ads performance response for cohort i can be represented as $w_i x_i + z_i$ and the simple linear programming problem can be formulated as follows:



(a) Cohort 0.



(b) Cohort 93.

Figure 3: Curve fitting of ads performance metric and engagement metric response.

$$\begin{aligned}
\max \sum_{0\le i<N-1} & k_i x_i + b_i \\
\text{s.t} \sum_{0\le i<N-1} & w_i x_i + z_i \ \geq \ (1-\rho)B, \\
& l \le x_i \le u, \ \forall i \in \{0, 1, ..., N-1\}.
\end{aligned} \tag{10}$$

In the above formulation (10), the objective engagement function is the sum of predicted engagement of all cohorts, the constraint ads performance function is also the sum of predicted ads performance of all cohorts. l and u are universal lower bound and upper bound for the scalar variable x_i. The advantages of LP formulation over ILP are: 1) the dimension of optimization variables is reduced from $N \times D$ to N, making it more scalable to multi-objective and multi-constraints formulation depending on business needs; 2) LP is easier to solve than ILP (P vs NP-hard); 3) the generated policy is continuous in stead of finite set; 4) linear fitting can handle certain outliers and thus yield a more generalizable result.

However, the LP formulation relies on the linearity assumption on ads performance metric and engagement metric, which may not be true for some cohorts; in addition, the user's sensitivity information for each cohort given by the MTML output is not utilized. In the next subsection, these side information are further incorporated into the optimization problem.

3.3 User-sensitivity-aware and Structure-aware Constraints

In the RCT dataset, we collected the ads performance metric and engagement metric response of each cohort for the Ad quality bid and the linearity of engagement response can be verified (see Figure 3a). However, there are also strong non-linear ads performance metric behavior in some other cohorts.

To better capture this non-linearity, quadratic fitting is applied to approximate the ads performance behavior for each cohort (see Figure 3b); as a result, the ads performance response of cohort i can be modeled as $a_i x_i^2 + b_i x_i + c_i$, where a_i, b_i and c_i are the coefficients of the quadratic forms. It is also worthwhile to note that linear fitting is a special case of quadratic fitting , hence the quadratic constraint is backward compatible for the cohorts whose ads performance metric behavior is linear.

In addition to the original LP formulation (10), we found it crucial to incorporate product constraints to enhance generalization and accelerate optimizer convergence. Users are categorized based on model-predicted sensitivities, and it is generally optimal to assign smaller scalars to more sensitive cohorts. To achieve this, we introduce a monotonic constraint: for cohort i and cohort j, $i, j \in \{0, 1, ..., N-1\}$, if $i \leq j$, then $x_i \leq x_j$, as cohorts are sorted in ascending order of sensitivity. This approach significantly improved generalization, as evidenced by enhanced performance on the test set, and also resulted in faster convergence.

The new linear programming problem formulation with quadratic constraint and monotonic constraint is finalized as follows:

$$
\begin{aligned}
\max \sum_{0 \leq i < N-1} & k_i x_i + b_i \\
\text{s.t} \sum_{0 \leq i < N-1} & a_i x_i^2 + b_i x_i + c_i \ \geq \ (1-\rho)B, \\
& l \leq x_i \leq u, \ \forall i \in \{0, 1, ..., N-1\}, \\
& x_i \leq x_j, \ \forall i \leq j \text{ and } i, j \in \{0, 1, ..., N-1\}.
\end{aligned} \tag{11}
$$

The final optimization problem can be solved by the embedded conic solver (ECOS) [Domahidi *et al.*, 2013], a classical primal-dual interior-point solver [Wright, 1997] for convex cone programs. ECOS reformulates our quadratically constrained problem into conic form by expressing the quadratic constraint as a second-order cone. It then solves the resulting second-order cone program (SOCP) using a primal-dual Newton method, iteratively updating primal and dual variables while maintaining feasibility within the cone. ECOS offers high numerical accuracy and is efficient for medium-scale problems, making it a practical choice for our use case.

4 Offline and Online Evaluations

In this section, offline and online evaluation results will be provided. The Opus framework is applied to two quality bids: quality bid 1 and quality bid 2, which use the prediction of a user's feedback toward an ad as the proxy for the ad's quality. In the MTML model, user cohorts for these two bids will be predicted simultaneously, then the predictions are fed into the optimization algorithm to obtain the personalized scalars individually (see Figure 2).

4.1 MTML Model Offline Performance

For the model architecture change from individual models to MTML (Figure. 2), AUCC (Area under Cost Curve) [Du *et al.*, 2019b] is used to measure business gain by a combination of ads performance and engagement. In an AUCC plot, the aggregated engagement gain is represented in the Y-axis and the aggregated ads performance loss is represented in the X-axis. The overall curve serves to help evaluate the return on investment predictions. The training and evaluation AUCC for the MTML model is shown in Figure 4. After MTML training is completed, a validation dataset is used to evaluate the performance of MTML model and the current model (baseline), the comparison of model performances is shown in Table 1. It can be observed that, with the new MTML architecture, there is a 2.4% AUCC increase in predicting user cohorts for quality bid 1 and a 1.9% AUCC increase in predicting user cohorts for quality bid 2. Therefore, the advantage of MTML architecture is clearly demonstrated.

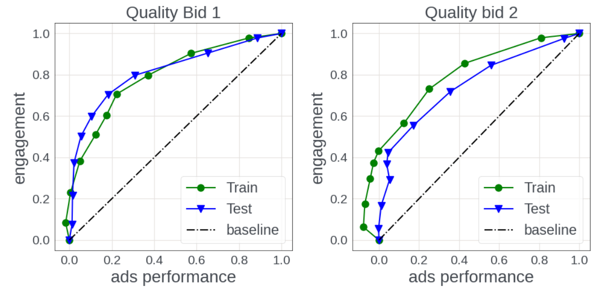


Figure 4: MTML model AUCC for two quality bids.

Table 1: Evaluation AUCC comparison between baseline models and MTML.

Eval AUCC	Quality bid 1	Quality bid 2
Baseline	0.748	0.792
MTML	**0.766** (+2.4%)	**0.807** (+1.9%)

4.2 Optimization Offline Performance

For the optimization part, two datasets containing each user cohort's ads performance metric and engagement metric responses at different time $t_0 < t_1$ are collected, denoted as D_{t_0}, D_{t_1}. In additions, each user is also randomly assigned a user bucket from 1 to 10 and the user bucket dimension can also be split into 2 disjoint sets $\{u_0, u_1\}$. The final training and testing dataset are D_{t_0,u_0} and D_{t_1,u_1} (see Figure 5). The reasons for creating disjoint training and testing dataset in both time and user bucket dimension are as follows:

- Split by time: avoid using future information to make prediction in the past and ensure that the model's performance is accurately evaluated when faced with future data.
- Split by user bucket: the dataset collected from RCT experiment is also a sampled version of production traffic. The user bucket split also helps prevent overfitting and boost generalizability of the policy derived from the optimization solution.

After solving the optimization problem and obtaining the scalar policy for each cohort, the predicted engagement metric and predicted ads performance metric for each cohort can also be computed by applying the scalar to the fitted function.

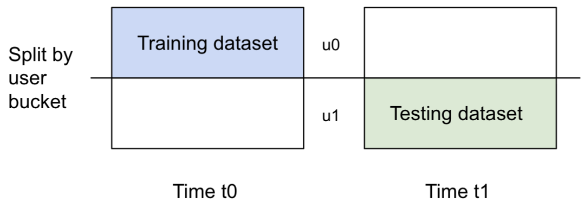


Figure 5: Training and testing dataset for optimization problem.

As a result, the summation of the predicted engagement metric and predicted ads performance metric for all cohorts will serve as the engagement proxy and ads performance proxy for the ads system. For better visualization, the absolute values of engagement and ads performance are not used; rather, they are compared with a baseline performance and the percentage of delta is computed. To be more concise, let (μ_0, θ_0) be the (ads performance metric,engagement metric) performance point evaluated under a uniform policy ($\alpha_i = 1$ for all cohorts i); denote (μ_1, θ_1) as the performance point evaluated under any other personalized scalar policy, the ads performance metric delta and engagement metric delta is defined as:

$$\text{ads performance metric delta} = \frac{\mu_1 - \mu_0}{\mu_0}, \tag{12}$$

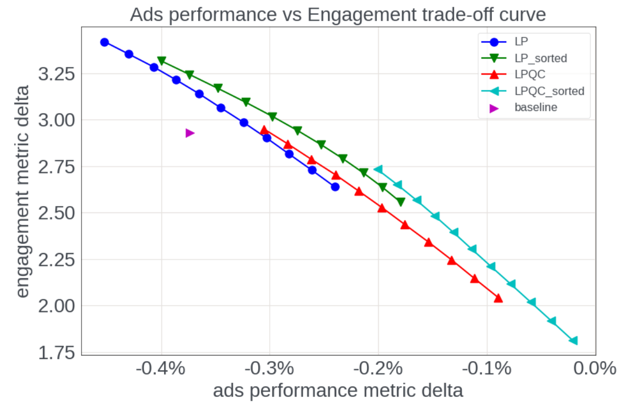
$$\text{engagement metric delta} = \frac{\theta_1 - \theta_0}{\theta_0}. \tag{13}$$

By changing the ads performance budget ρ, which represents the business needs, an ads performance vs engagement trade-off curve can be generated to show the performance or efficiency of the optimization approach. In this paper, we generated and compare the performance of the following 4 methods:

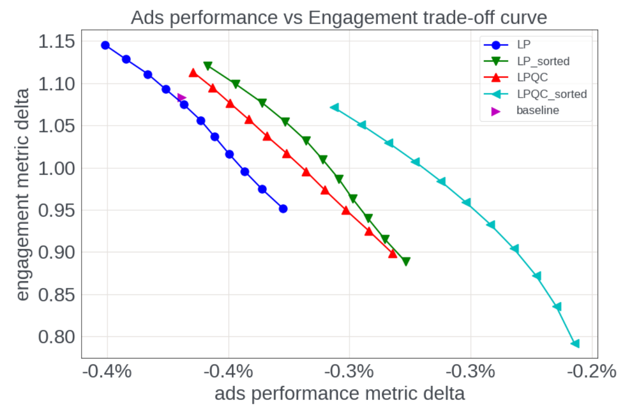
- LP: linear programming method defined in (10).
- LP_sorted: linear programming method plus monotonic constraint.
- LPQC: linear programming method with quadratic constraint in revneue, *i.e.* replacing linear modeling of ads performance metric in (10) with quadratic modeling.
- LPQC_sorted: linear programming method with quadratic and monotonic constraint, as defined in (11).

In addition to the above 4 performance curves, an isolated ads performance vs engagement performance point under current baseline's scalar policy can also be evaluated. Current baseline's policy is obtained by first predicting user sensitivity with a single deep neural network model and then solving the linear programming problem in (10). The overall performance comparison for quality bid 1 is shown in Figure 6a and several highlights can be observed:

- 4 trade-off curves are located to the top-right area of the current baseline's performance point, which means that under certain scalar policies, ads performance metric gain and engagement metric can be achieved simultaneously.



(a) Quality bid 1.

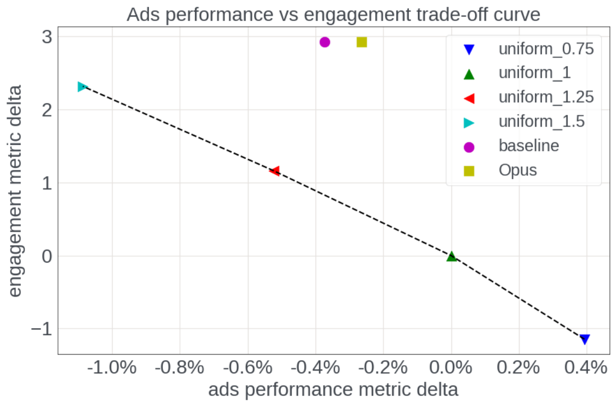


(b) Quality bid 2.

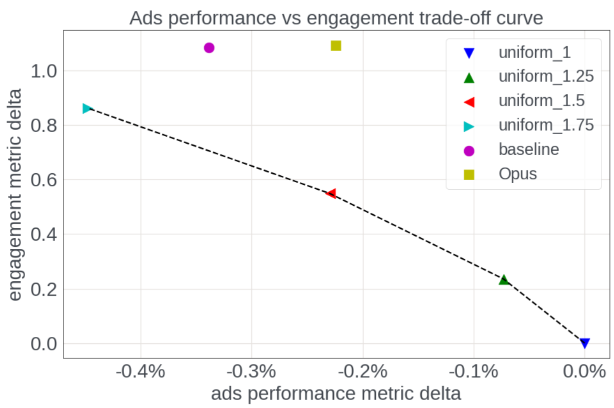
Figure 6: Ads performance vs engagement trade-off curves of different methods for quality bid 1 and 2.

- Under the same level of engagement, there is about 0.06% offline ads performance metric delta gain from current baseline's performance point to the LP performance curve; given that the only difference between them is the user sensitivity prediction part (single DNN model vs MTML), the offline ads performance metric improvement further validates the advantage of MTML model.
- Comparing LP vs LPQC and LP_sorted vs LPQC_sorted, it can be observed that by changing the ads performance metric modeling from linear to quadratic, the ads performance vs engagement performance curve shifts to the right by about 0.01% to 0.02% ads performance delta gain, demonstrating that quadratic modeling better captures the ads performance metric behavior and achieve ads performance gain.
- Comparing LP vs LP_sorted and LPQC vs LPQC_sorted, it can also be observed that by adding the monotonic constraint, there is roughly 0.04% to 0.05% positive ads performance metric delta shift of the performance curve, which is a clear indication of the importance in user sensitivity information in the optimization problem.

Similar results can be derived for quality bid 2 (Figure. 6b) and they are omitted here due to space constraint. With the



(a) Quality bid 1.



(b) Quality bid 2.

Figure 7: Ads performance vs engagement trade-off curves of uniform policy vs performance from Opus for quality bid 1 and 2.

performance curves and the current baseline's performance point, it is easy to fine-tune the ads performance budget ρ to find the scalar policy that achieves the business need. For example, if the business objective is to increase ads performance metric with neutral engagement; a horizon line can be drawn across the current baseline's performance point and the intersection of the horizontal line with the performance curve will be the solution that maximizes ads performance metric; and the scalar policy associated with the intersection point will be the optimal policy. Similarly, a vertical line can be drawn if the goal is to increase engagement without ads performance metric loss. It is also worthwhile to compare the personalization result with uniform policies. Figure 7a shows the performance trade-off curve when applying uniform policies under different scalars for quality bid 1, as the scalar gets larger, the systems transition from a high ads performance but low engagement to region into a low ads performance but high engagement region. However, the system's performance is limited by the performance curve; personalization enables the system to explore area with higher ads performance and engagement. By simple DNN model for user sensitivity prediction and LP formulation (10), the system achieved 3 times engagment metric delta gain with roughly

Table 2: Comparison of normalized efficiency between current baseline and Opus.

Metrics	Normalized efficiency
Baseline	13.96%
Opus	$\mathbf{20.12}\%$

Table 3: Online performance of Opus framework.

Metrics	Ads performance	DAU	Time-spent
Baseline	100%	100%	100%
Pretest (Opus)	$\mathbf{100.5}\%$	99.999%	100.0072%
Backtest (Opus)	$\mathbf{100.47}\%$	99.996%	100.0085%

0.4% ads performance metric spent compared to baseline policy; the Opus framework further increase the ads performance metric by roughly 0.15% with no engagement metric loss, as a result, from the ads performance metric's perspective, the Opus method is equivalent to applying a uniform scalar (roughly 1.1) but from the engagement metric's perspective, it achieves more than 3 times engagement metric delta gain than the uniform scalar policy. Apart from focusing on individual ads performance or engagement metric gain while fixing the other metric, the ads performance over engagement efficiency metric provides another perspective to quantify the power of the algorithm. The efficiency metric is defined as the product of some normalization constants C_0 and the ratio of engagement metric delta over ads performance metric delta, which are already defined in (12):

$$\begin{aligned} \text{efficiency} &= C_0 \cdot \frac{\text{engagement metric delta}}{\text{ads performance metric delta}} \\ &= C_0 \cdot \frac{\theta_1 - \theta_0}{\mu_1 - \mu_0}. \end{aligned} \tag{14}$$

With the offline dataset, the efficiency of current baseline's policy and the proposed Opus framework can be computed; it can be observed from Table 2 that the Opus framework achieves **1.44x** efficiency boost compared to current baseline.

4.3 Online Results

With the promising offline result, an online A/B test is conducted (one-month) before the launch and a backtest is also conducted (one-month). Our current business goal is to maximize ads performance without engagement loss, hence according to Figure 6, a horizontal line is drawn and intersect with the LPQC_sorted curve; the scalar policy associated with the intersection point will be the personalized policy for online testing. Similar steps are also applied to the quality bid 2 to obtain a new scalar policy for each cohort.

After applying the new policy treatment, the overall online ads performance and engagement performance is shown in Table 3. With the new Opus framework, around 0.5% ads performance metric gain with neutral engagement metric can be achieved.

5 Conclusions and Future Works

In this paper, a personalization framework call Opus is developed, aiming to customize ad's quality bid scalar for different user cohorts and can achieve both ads performance metric and engagement metric gain. With the MTML model, user sensitivity is first predicted and later linear programming problem with quadratic and monotonic constraint is formulated to solve for the scalar policy. The new policy achieves around 0.5% ads performance metric gain in online experiments, as well as 1.44x efficiency boost. For future directions, we are planning to extend the user sensitivity prediction from single snapshot model prediction to dynamic user sensitivity prediction; we are also planning to extend to optimization formulation to multi-objective, multi-constraint optimization.

6 GenAI Usage Disclosure

In accordance with the ACM's Authorship Policy, we provide the following disclosure regarding the use of Generative AI (GenAI) tools in the preparation of this research paper:

- During the research and development stage, no GenAI tool was utilized to assist in the collection, processing and analysis of large datasets, nor was it used in the codebase.
- During the draft writing and editing, no GenAI tool was used to help initial text for any sections, nor was it used for refining of the draft.

References

[Agarwal *et al.*, 2014] Deepak Agarwal, Souvik Ghosh, Kai Wei, and Siyu You. Budget pacing for targeted online advertisements at linkedin. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 1613–1619, 2014.

[Agarwal *et al.*, 2018] Deepak Agarwal, Kinjal Basu, Souvik Ghosh, Ying Xuan, Yang Yang, and Liang Zhang. Online parameter selection for web-based ranking problems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 23–32, 2018.

[Alaa *et al.*, 2023] Ahmed M. Alaa, Zaid Ahmad, and Mark van der Laan. Conformal meta-learners for predictive inference of individual treatment effects. In *NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 47682–47703, 2023.

[Boyd and Vandenberghe, 2004] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, Massachusetts, 2004.

[Cai *et al.*, 2017] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. Real-time bidding by reinforcement learning in display advertising. In *WSDM '17: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 661–670, 2017.

[Domahidi *et al.*, 2013] Andreas Domahidi, Eric Chu, and Stephen Boyd. Ecos: An socp solver for embedded systems. *European Control Conference (ECC)*, pages 3071–3076, 2013.

[Du *et al.*, 2019a] Ruihuan Du, Yu Zhong, Harikesh Nair, Bo Cui, and Ruyang Shou. Causally driven incremental multi touch attribution using a recurrent neural network. *arXiv preprint arXiv:1902.00215*, 2019.

[Du *et al.*, 2019b] Shuyang Du, James Lee, and Farzin Ghaffarizadeh. Improve user retention with causal learning. In *The 2019 ACM SIGKDD Workshop on Causal Discovery.*, pages 34 – 39. PMLR, 2019.

[Gao *et al.*, 2024] Jingtong Gao, Bo Chen, Menghui Zhu, Xiangyu Zhao, Xiaopeng Li, Yuhao Wang, Yichao Wang, Huifeng Guo, and Ruiming Tang. Hierrec: Scenario-aware hierarchical modeling for multi-scenario recommendations. In *CIKM '24: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 653–662, 2024.

[Jeunen *et al.*, 2024] Olivier Jeunen, Jatin Mandav, Ivan Potapov, Nakul Agarwal, Sourabh Vaid, Wenzhe Shi, and Aleksei Ustimenko. Multi-objective recommendation via multivariate policy learning. In *Proceedings of the 18th ACM Conference on Recommender Systems*, page 712–721, 2024.

[Jr. *et al.*, 2013] David W. Hosmer Jr., Stanley Lemeshowm, and Rodney X. Sturdivant. *Applied Logistic Regression*. John Wiley & Sons, Inc, 2013.

[Kalra *et al.*, 2023] Achir Kalra, Chong Wang, Cristian Borcea, and Yi Chen. Reserve price optimization in first-price auctions via multi-task learning. In *2023 IEEE International Conference on Data Mining (ICDM)*, pages 200–209, 2023.

[Künzel *et al.*, 2019] Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. In *Proceedings of the national academy of sciences*, page 4156–4165, 2019.

[Louizos *et al.*, 2017] Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6449–6459, 2017.

[Ma *et al.*, 2018a] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1930–1939, 2018.

[Ma *et al.*, 2018b] Xiao Ma, Liqin Zhao, Guan Huang, Zhi Wang, Zelin Hu, Xiaoqiang Zhu, and Kun Gai. Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1137–1140, 2018.

[Ma *et al.*, 2022] Ning Ma, Mustafa Ispir, Yuan Li, Yongpeng Yang, Zhe Chen, Derek Zhiyuan Cheng, Lan Nie, and Kishor Barman. An online multi-task learning framework for google feed ads auction models. In *KDD '22: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3477–3485, 2022.

[MacQueen and B., 1967] MacQueen and J. B. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[Radcliffe and Surry, 2011] Nicholas J Radcliffe and Patrick D Surry. Real-world uplift modelling with significance-based uplift trees. *White Paper TR-2011-1, Stochastic Solutions*, pages 1—33, 2011.

[Raul *et al.*, 2023] Ameya Raul, Amey Porobo Dharwadker, and Brad Schumitsch. Cam2: Conformity-aware multi-task ranking model for large-scale recommender systems. In *WWW '23 Companion: Companion Proceedings of the ACM Web Conference 2023*, pages 513–517, 2023.

[Rubin, 2005] Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.*, 100(469):322–331, 2005.

[Saito *et al.*, 2020] Yuta Saito, Hayato Sakata, and Kazuhide Nakata. Cost-effective and stable policy optimization algorithm for uplift modeling with multiple treatments. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, page 406–414, 2020.

[Shalit *et al.*, 2017] Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms. In *ICML'17: Proceedings of the 34th International Conference on Machine Learning*, pages 3076–3085, 2017.

[Shi *et al.*, 2024] Wei Shi, Chen Fu, Qi Xu, Sanjian Chen, Jizhe Zhang, Qinqin Zhu, Zhigang Hua, and Shuang Yang. Ads supply personalization via doubly robust learning. In *CIKM '24: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 4874–4881, 2024.

[Tang *et al.*, 2020] Hongyan Tang, Junning Liu, Ming Zhao, and Xudong Gong. Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In *RecSys '20: Proceedings of the 14th ACM Conference on Recommender Systems*, pages 269–278, 2020.

[Tang *et al.*, 2024] Jie Tang, Huiji Gao, Liwei He, and Sanjeev Katariya. Multi-objective recommendation via multivariate policy learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, page 5783–5792, 2024.

[Wright, 1997] Stephen J Wright. *Primal-dual interior-point methods*. SIAM, 1997.

[Zhang *et al.*, 2014] Weinan Zhang, Shuai Yuan, and Jun Wang. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 1077–1086, 2014.

[Zhang *et al.*, 2020] Wenhao Zhang, Wentian Bao, Xiao-Yang Liu, Keping Yang, Quan Lin, Hong Wen, and Ramin Ramezani. Large-scale causal approaches to debiasing post-click conversion rate estimation with multi-task learning. In *WWW '20: Proceedings of The Web Conference 2020*, pages 2775–2781, 2020.

[Zhao *et al.*, 2017] Yan Zhao, Xiao Fang, and David Simchi-Levi. Uplift modeling with multiple treatments and general response types. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, page 588–596, 2017.

[Zhao *et al.*, 2020] Xiangyu Zhao, Xudong Zheng, Xiwang Yang, Xiaobing Liu, and Jiliang Tang. Jointly learning to recommend and advertise. In *KDD '20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3319–3327, 2020.