NOAH

Neural **O**ptimization with **A**daptive **H**euristics for Intelligent Marketing System

Changshuai Wei

Changshuai Wei*, Benjamin Zelditch, Joyce Chen, Andre Assuncao Silva T Ribeiro, Kenneth Tay, Borja Ocejo Elizondo, Sathiya Keerthi Selvaraj, Aman Gupta, Licurgo Benemann De Almeida

*Corresponding author

LinkedIn

Outline

1. Challenges
2. Framework
3. Application
4. Learnings

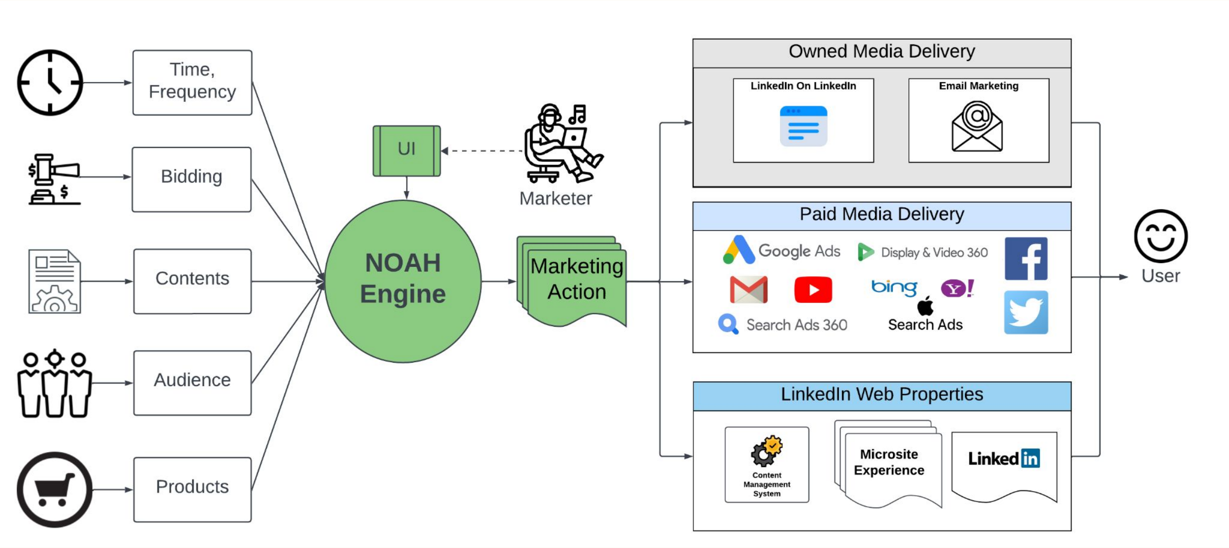
Marketing AI System: Challenges

- Heterogeneous Data
  - Owned channel and paid channel
- Complex action space
  - Audience, content, bidding, timing, etc
- Multiple business goals and constraints
  - 2B products and 2C products
  - Revenue and user experience

NOAH Framework

First General Framework that considers:

- Both 2B and 2C products
- Both owned and paid channels

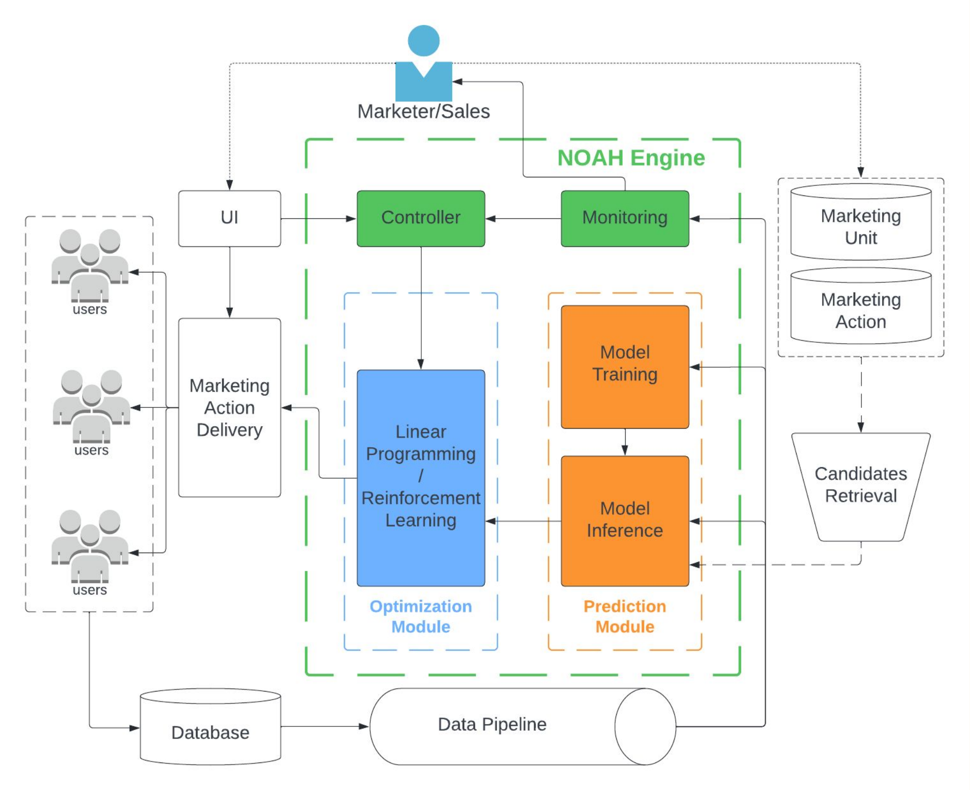


Linked in

NOAH Framework

Core Components

- **N**eural Module
  - MTL
  - Causal ML
- **O**ptimization Module
  - Large-scale LP
  - RL
- **A**daptive **H**euristics
  - Controller
  - Agent

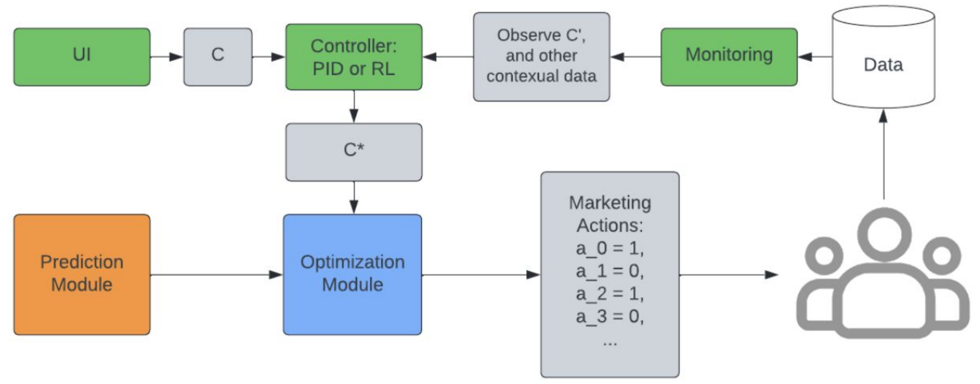


NOAH Framework

$$y = f(x_u, x_a, x_c),$$

$$\begin{aligned} \max_{a_u} & \sum_u f_0(x_u, x_{a_u}, x_c), \\ \text{s.t.} & \sum_u f_k(x_u, x_{a_u}, x_c) \leq C_k, \end{aligned}$$

- Guidelines
  - Multiple actions: sequential
  - Optimization strategy: LP vs RL
- Common Pattern
  - Causal ML + LP + Controller



NOAH Algorithm: Bidding Optimization

$$\max_{p(u)} \sum_u y^{\text{rev}}_{u,p(u)},$$
$$\text{s.t.} \frac{\sum_u y^{\text{rev}}_{u,p(u)}}{\sum_u y^{\text{cost}}_{u,p(u)}} \geq C_{ROAS},$$

$$g(\lambda) = \max_{p(u)\in\{1,\dots,P\}} \sum_u y^{\text{rev}}_{u,p(u)} - \lambda \left(C_{ROAS} \sum_u y^{\text{cost}}_{u,p(u)} - \sum_u y^{\text{rev}}_{u,p(u)} \right)$$
$$= \sum_u \max_{p(u)\in\{1,\dots,P\}} \left[(1+\lambda) y^{\text{rev}}_{u,p} - \lambda C_{ROAS} y^{\text{cost}}_{u,p} \right]$$

$$\partial g(\lambda)|_{\lambda=\lambda'} = \sum_u y^{\text{rev}}_{u,p_{\lambda'}(u)} - C_{ROAS} \sum_u y^{\text{cost}}_{u,p_{\lambda'}(u)}$$

Algorithm 1 Bidding Optimization with Lagrangian Relaxation

Input: $y^{\text{rev}}_{u,p}$, $y^{\text{cost}}_{u,p}$, C_{ROAS}, $\lambda_{\text{init}} > 0$, and $0 < \epsilon \ll \lambda_{\text{init}}$
Output: $p^*(u)$ for each u

- set $\lambda_{min} \leftarrow 0$ and $\lambda_{max} \leftarrow \lambda_{\text{init}}$
- **if** $\partial g(\lambda)|_{\lambda=0} \geq 0$ **then**
  - **return** $p_0(u)$
- **else**
  - **while** $\partial g(\lambda)|_{\lambda=\lambda_{max}} < 0$ **do**
    - set $\lambda_{min} \leftarrow \lambda_{max}$ and $\lambda_{max} \leftarrow 2\lambda_{max}$
  - **end while**
- **end if**
- **while** $\lambda_{max} - \lambda_{min} > \epsilon$ **do**
  - set $\lambda' \leftarrow \frac{\lambda_{min}+\lambda_{max}}{2}$
  - **if** $\partial g(\lambda)|_{\lambda=\lambda'} > 0$ **then**
    - set $\lambda_{max} \leftarrow \lambda'$
  - **else if** $\partial g(\lambda)|_{\lambda=\lambda'} < 0$ **then**
    - set $\lambda_{min} \leftarrow \lambda'$
  - **else**
    - **break**
  - **end if**
- **end while**
- **return** $p_{\lambda'}(u)$

NOAH Algorithm: Content Optimization

$$\rho^{-1}(y) = x^T w + \epsilon, \qquad \epsilon \sim N(0, \beta^2),$$

- Online learning
- High-dimensional setting
- On-device computing
- Assumption: independent gaussian distributed weights

Algorithm 2 Content Optimization with Online Contextual Bandit

Input: $\mu_i, \sigma_i^2, \beta^2$

for time step $j = 1, 2, \cdots, J$ **do**

Sample $\hat{w}_i$ from $N(\mu_{w_i}, \sigma_{w_i}^2)$ **for all** i;

for creative $a = 1, 2, \cdots, A$ **do**

Sample $\hat{\epsilon}_a$ from $N(0, \beta^2)$

Calculate latent reward $\tilde{t}_{j,a} = \hat{\epsilon}_a + \sum_i x_{i,j,a}\hat{w}_i$

end for

Choose creative $a_j = \arg\max_a \tilde{t}_{j,a}$ and show it to user; observe user engagement and receive reward y_j

Update $\mu_{w_i} \leftarrow \mu_{w_i} + x_i \frac{\sigma_{w_i}^2}{\tau^2}\Delta$;

Update $\sigma_{w_i}^2 \leftarrow \sigma_{w_i}^2 [1 - x_i^2 \frac{\sigma_{w_i}^2}{\tau^2}(1 - \gamma)]$

end for

Here, the value of γ and Δ depends on the link function. For linear regression, $\gamma = 0$ and $\Delta = y - x^T w$. For Probit regression, $\gamma = 1 - \varpi(\frac{y\eta}{\tau})$ and $\Delta = y\tau\vartheta(\frac{y\eta}{\tau})$, where, $\vartheta(t) = \frac{\mathcal{N}(t;0,1)}{\Phi(t;0,1)}$ and $\varpi(t) = \vartheta(t)(\vartheta(t) + t)$.

NOAH Application: Email marketing

Major A/B test win: one of largest in LinkedIn algorithmic marketing thus far

Table 1: Overview of A/B Test Result

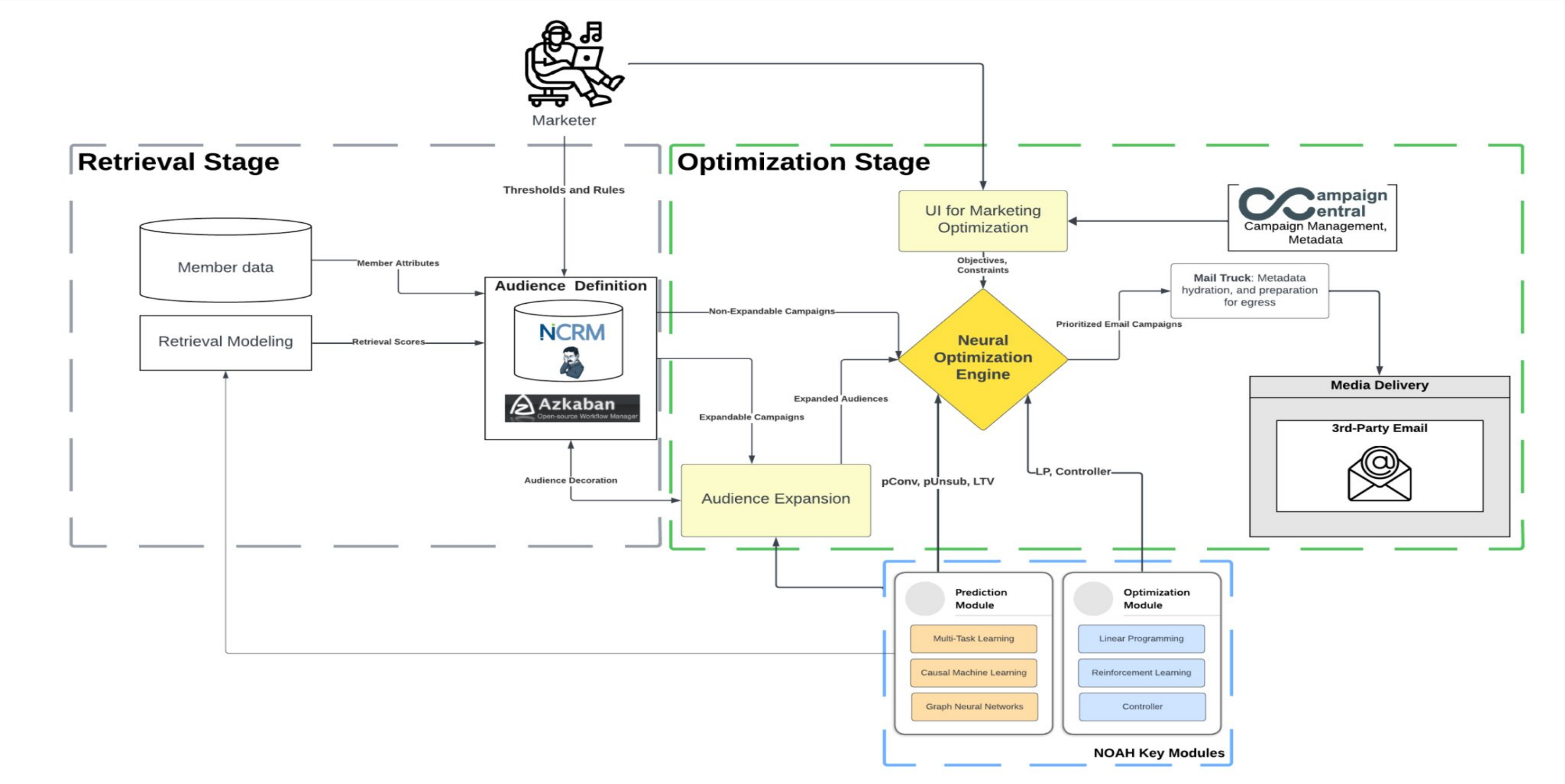
Marketing System	Surrogate Index	Unsub	Email Volume
Legacy+AE	neutral	+8.08%	+9.42%
NOAH	+2.16%[1]	+1.52%	+8.44%

[1] All percentages are relative to the control arm.

Product Type	Surrogate Index	3-Mth Metric[1]	Conv
2C	not stat sig	neutral	neutral
2B	+13.21%	+23.66%	+13.30%

[1] Observed 3-month long-term metric q^3 after test concluded.

Email System



Large scale LP with randomization: original problem

$$
\begin{aligned}
\min_{a_{u,j}} \; & \sum_{u,j} -y^{\text{conv}}_{u,j}\, y^{\text{ltv}}_{u,p(j)}\, a_{u,j}, \\
\text{s.t.} \; & \sum_{u,j} a_{u,j} y^{\text{unsub}}_{u,j} \leq C_{\text{unsub}}, \\
& \sum_{u} \sum_{j \in \mathbb{J}_{2B}} -a_{u,j} \leq -C_{2B}, \\
& \sum_{u} \sum_{j \in \mathbb{J}_{2C}} -a_{u,j} \leq -C_{2C}, \\
& \sum_{j} a_{u,j} \leq C_{\text{fcap}}, \quad \forall u, \\
& a_{u,j} \in \{0, 1\}, \quad \forall u, j.
\end{aligned}
$$

Large scale LP with randomization: algorithm

Original problem

MIP to LP, add regularization

$$\min_a \; y^\top a + \frac{\gamma}{2} a^\top a \quad s.t. \quad Da \le b, \quad a_u \in \mathcal{B}_u,$$

Dual problem

$$g_\gamma(\lambda) = \min_{a \in \mathcal{B}} \left\{ y^\top a + \frac{\gamma}{2} a^\top a + \lambda^\top (Da - b) \right\},$$

Main iterations

$$\nabla g_\gamma(\lambda) = D\hat{a}(\lambda) - b,$$

$$\hat{a}_u(\lambda) = \Pi_{\mathcal{B}_u}\left[-\frac{1}{\gamma}(D^\top \lambda + y)_u \right],$$

Sampling primal

(1) Calculate sampling probabilities $p_{u,j} = \frac{a_{u,j}}{\sum_j a_{u,j}}$

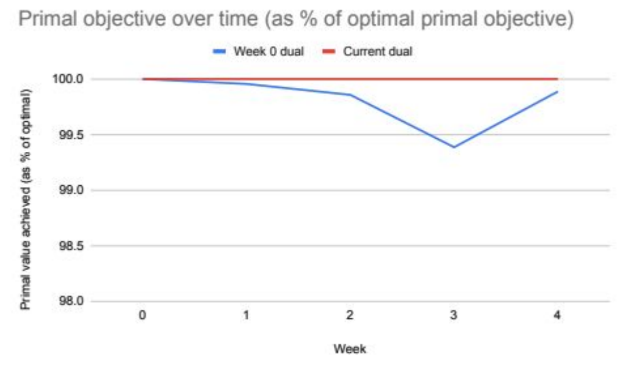
(2) Sample a set $\mathcal{J}'_u$ of size $\lfloor \sum_j a_{u,j} + 0.5 \rfloor$ without replacement, from set of all index $\mathcal{J}_u$ with non-uniform distribution of $p_{u,j}$

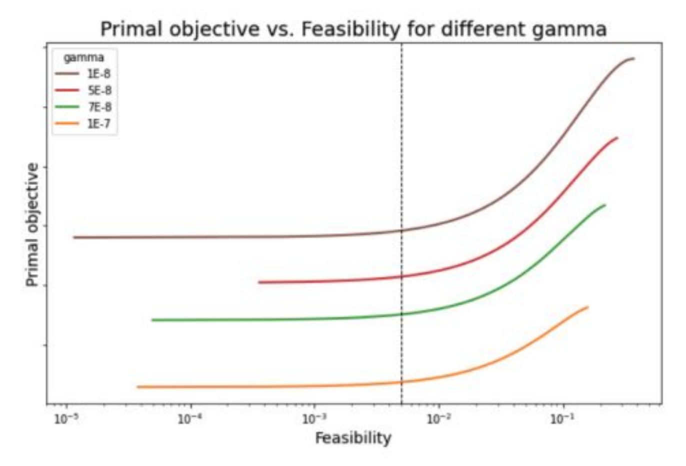
(3) We obtain $a'_{u,j} = I(j \in \mathcal{J}'_u)$, where $I(\cdot)$ is indicator function.

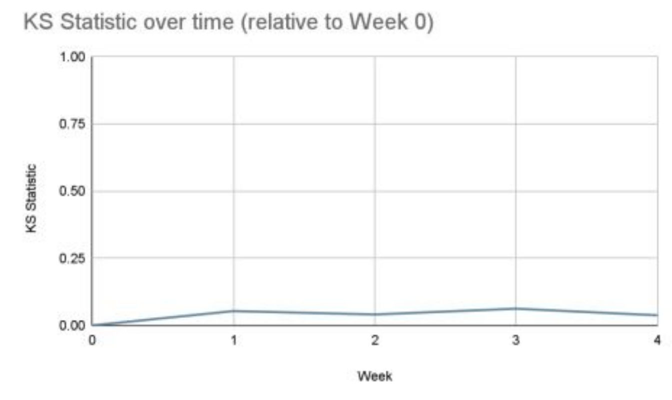
- MIP relaxation to LP
- Solving large scale LP
  - LP to QP with regularization
  - Parallelize primal update
  - Efficient box-cut projection (Wolfe's algo)
- Main iterations
  - gradient descent to update dual
  - box-cut projection to update primal

Large scale LP with randomization: learnings

- Smaller gamma gives better optimality
- Majority of primal are integer solution (on average 95%)
- Use of previous dual for current primal solution







Linkedin

LTV modeling for delayed feedback

$$y_{u,p}^{\mathrm{ltv}} = f_{\mathrm{ltv}}(x_u, x_p, x_c)$$

$$L = \sum_i \left[\mu_i k + q_i^{12} \exp(-\mu_i) k \right]$$

- Using LTV to predict long term metrics
- Handling of censored observations
    - Gamma model + axillary tasks
    - Survival model
- LTV used in optimization
- LTV used in AB test

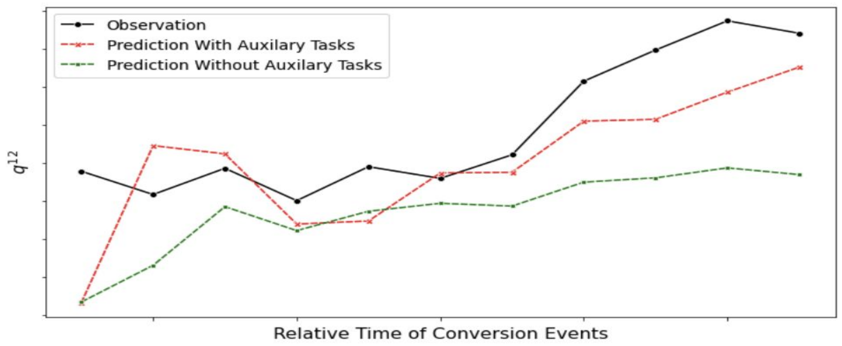


Figure 9: Gamma model with and without auxiliary tasks

Table 6: Spearman rank correlation of Survival and Gamma Models

Product Group	Survival	Gamma	Rel. Diff.
2C #1	0.448	0.491	+9.6%
2C #2	0.526	0.557	+5.9%
2C #3	0.425	0.489	+15%
2B #1	0.336	0.369	+9.8%
2B #2	0.458	0.396	-13%

LinkedIn

Audience Expansion to improve Retrieval

- Conservative audience list in legacy system
  - To avoid “overlap” due to lack of optimization stage
- ANN to expand conservative audience list
  - Locally sensitive hashing
- Trade-off: audience size and run-time

$$\mathcal{U}_j^{\text{expansion}} \subseteq \left\{ u \in \mathcal{U}_j^- : \exists u' \in \mathcal{U}_j^{\circ}, \quad \text{s.t. } \|x_u - x_{u'}\|_2 \leq \delta \right\}$$

δ	Audience Size	LP Output Size	Overall Runtime
0.00001	+0.38%	+0.33%	+4.21%
0.00005	+1.04%	+0.95%	+4.77%
0.0001	+3.77%	+3.23%	+15.42%
0.0005	+28.42%	+8.44%	+16.11%
0.001	>+250%	+9.59%	>+300%

AB test Design to reduce signal Dilution

- Reduce Dilution
  - Restrict to member who are “available” to receive marketing email
  - Conversion can happen on “non-exposed” audience
- Counterfactual comparison on AE and non-AE arm

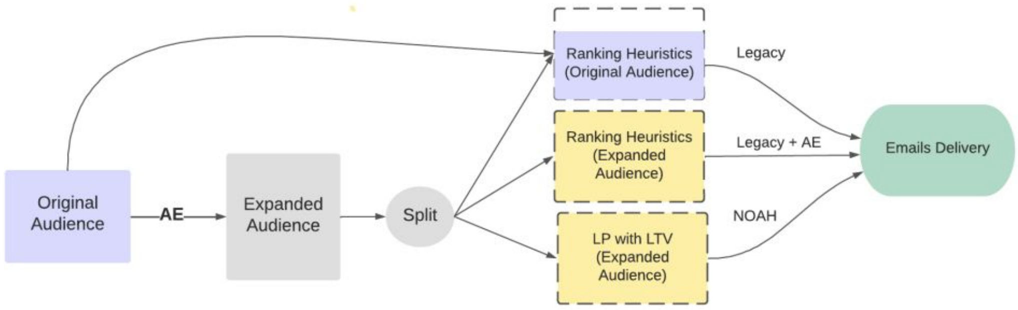


Figure 3: AB Test Flow

Statistical testing for zero-inflated heavy tail metrics

- Surrogate Index for long term value
  - Zero-inflated heavy tail metrics
- T-test robustness
  - Simulated log-normal distribution
  - Sampled distribution from SI
- Alternatives that doesn't work
  - Winsorization
  - Non-para test

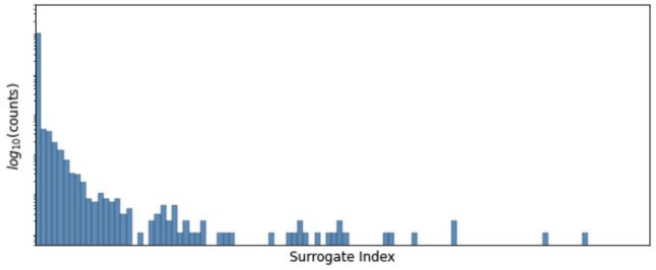


Figure 5: Distribution of Surrogate Index

Table 3: Type 1 errors of t test against zero-inflated heavy-tail distribution

α[1]	$X' \sim P_{obs}$		$X' \sim P_{ln}$	
	$p_0 = 0.02$	$p_0 = 0.8$	$p_0 = 0.02$	$p_0 = 0.8$
0.05	0.03038	0.04837	0.02464	0.03917
0.01	0.00227	0.00858	0.00168	0.00524
0.005	0.00072	0.00376	0.00048	0.00211
0.001	0.00002	0.0005	0.00002	0.00031

[1] Size of the test.

Thank you