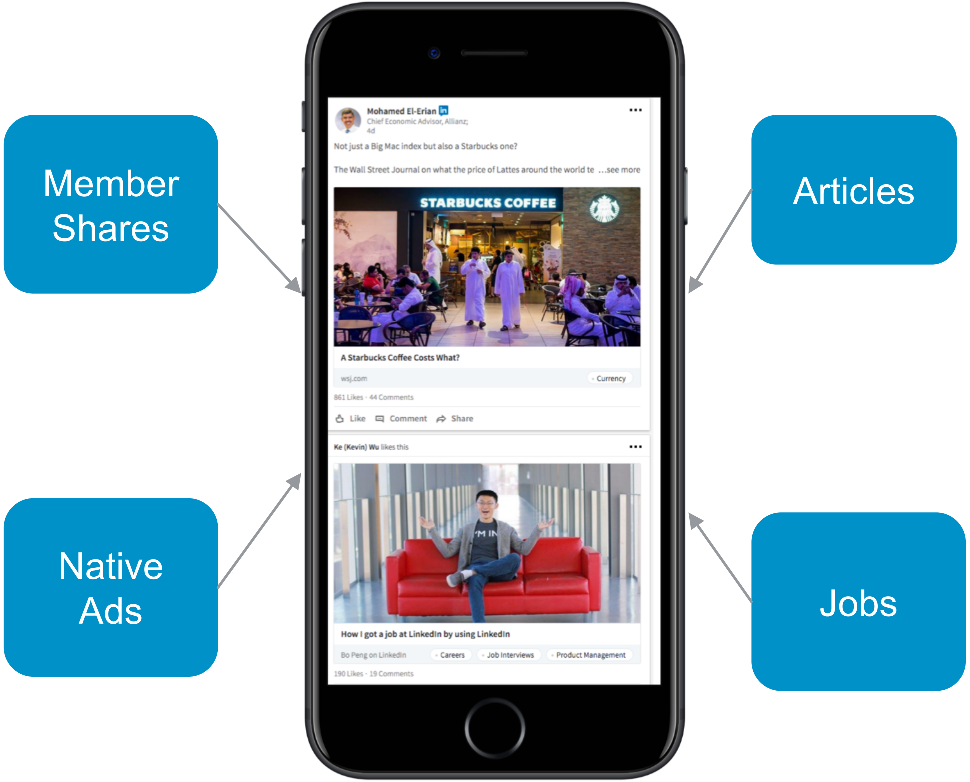
Online Parameter Selection for Web-based Ranking

Deepak Agarwal

Joint Work with Kinjal Basu, Souvik Ghosh, Ying Xuan, Yang Yang and Liang Zhang.

LinkedIn Feed



Mission: Enable Members to build an active professional community that advances their career.

The Feed is:

- The personalized “home page” of LinkedIn
- A heterogenous list of updates
  - Shares from a member’s connections
  - Recommendations including jobs, articles, connections, courses
  - Sponsored content or ads.

Important Metrics

- **Viral Action (VA)** - Sessions where members liked, shared or commented on an item
- **Job Applies (JA)** - Sessions where members applied for a job
- **Engaged Feed Session (EFS)**- Sessions where a member engaged with anything on feed.

Ranking

$$S(m,u) := P_{VA}(m,u) + x_{EFS}P_{EFS}(m,u) + x_{JA}P_{JA}(m,u).$$

- The weight vector $\boldsymbol{x} = (x_{EFS}, x_{JA})$ controls the balance of the metrics EFS, VA and JA.
- Business Strategy:

$$\begin{array}{ll} \underset{\boldsymbol{x} \in \mathcal{X}}{\text{Maximize}} & VA(\boldsymbol{x}) \\ \text{subject to} & EFS(\boldsymbol{x}) \geq c_{EFS}, JA(\boldsymbol{x}) \geq c_{JA} \end{array}$$

Problem

- The optimal value of x (tuning parameters) changes over time
- Example of changes
  - New content types are added
  - Relevance models are updated

- With every change engineers would manually find the optimal x
  - Run multiple A/B tests
- Not the best use of engineering time

Metrics Modeling

- $Y^k_{i,j}(\boldsymbol{x}) \in \{0, 1\}$ denote if the the i-th member during the j-th session which was served by parameter x, did action k or not. Here k = VA, EFS or Job Clicks.

- We model this data as follows

$$Y^k_i(\boldsymbol{x}) \sim Binomial(n_i(\boldsymbol{x}), \sigma(f_k(\boldsymbol{x}))),$$

where $n_i(x)$ is the total number of sessions of member i which was served by x and f_k is a latent function for the particular metric.

- Assume a Gaussian process prior on the latent function f.

Reformulation

We estimate the metrics as

$$VA(\boldsymbol{x}) = \sigma(f_{VA}(\boldsymbol{x}))$$
$$EFS(\boldsymbol{x}) = \sigma(f_{EFS}(\boldsymbol{x}))$$
$$JA(\boldsymbol{x}) = \sigma(f_{JA}(\boldsymbol{x}))$$

Solve the unconstrained optimization problem:

$$\underset{x \in \mathcal{X}}{\text{Maximize}}\ VA(\boldsymbol{x}) + \lambda(\sigma_{\xi}(EFS(\boldsymbol{x}) - c_{EFS}) + \sigma_{\xi}(JA(\boldsymbol{x}) - c_{JA}))$$

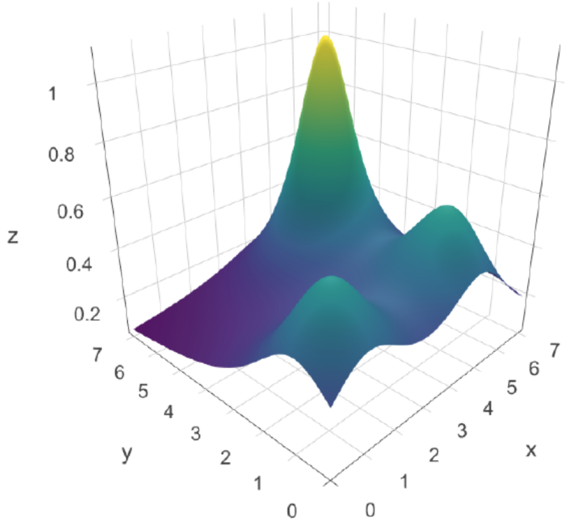
Benefit: Equivalent to finding the maximum of a continuous functional of a Gaussian process

Thompson Sampling Algorithm

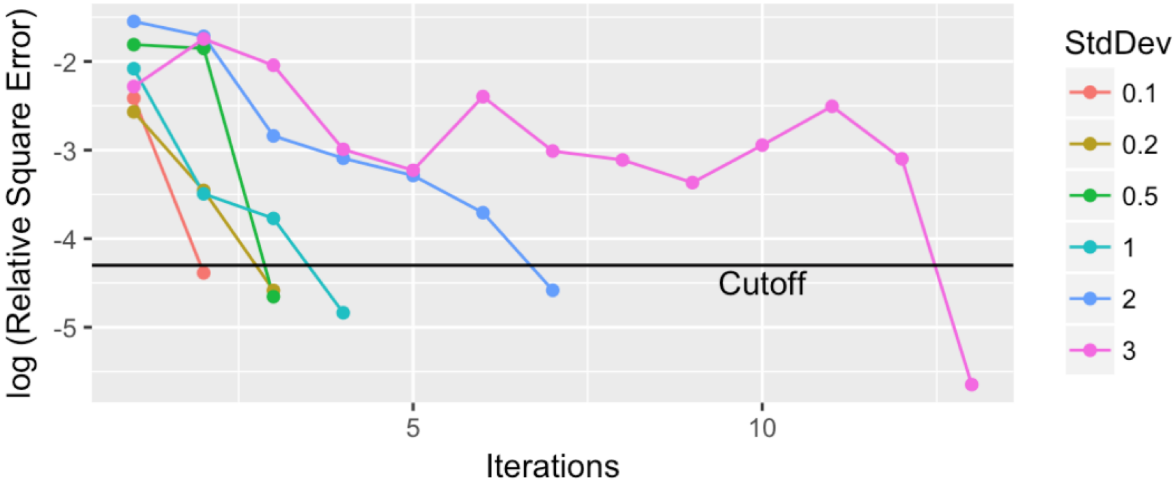
- Consider a Gaussian Process Prior on each f_k, where k is VA, EFS or JA
- Observe the data $Y_i^k(\boldsymbol{x})$
- Obtain the posterior of each f_k which is another Gaussian Process
- Sample from the posterior distribution and generate samples for the overall objective function.
- We get the next distribution of hyperparameters by maximizing the sampled objectives (over a grid of QMC points).

Continue until the distribution of x converges.

Simulation Results



(a) Trimodal Shekel Function



(b) Decay of log relative square error

Online Serving

- The Thompson Sampling approach generates a sampling distribution at every iteration.

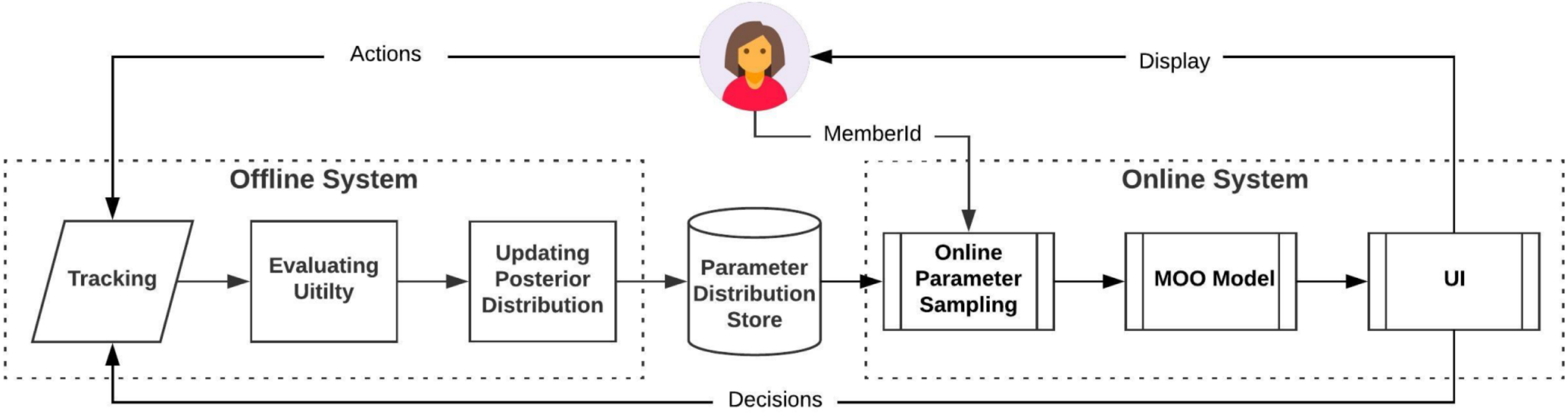
$$(x_1, \ldots, x_n) \text{ with probability } (p_1, \ldots, p_n) \text{ such that } \sum p_i = 1$$

- To serve members with the same distribution, each memberId is mapped to [0,1] using a hashing function h. For example, if

$$\sum_{i=1}^{k} p_i < h(Deepak) \leq \sum_{i=1}^{k+1} p_i$$

Then the feed is served with parameter x_{k+1}

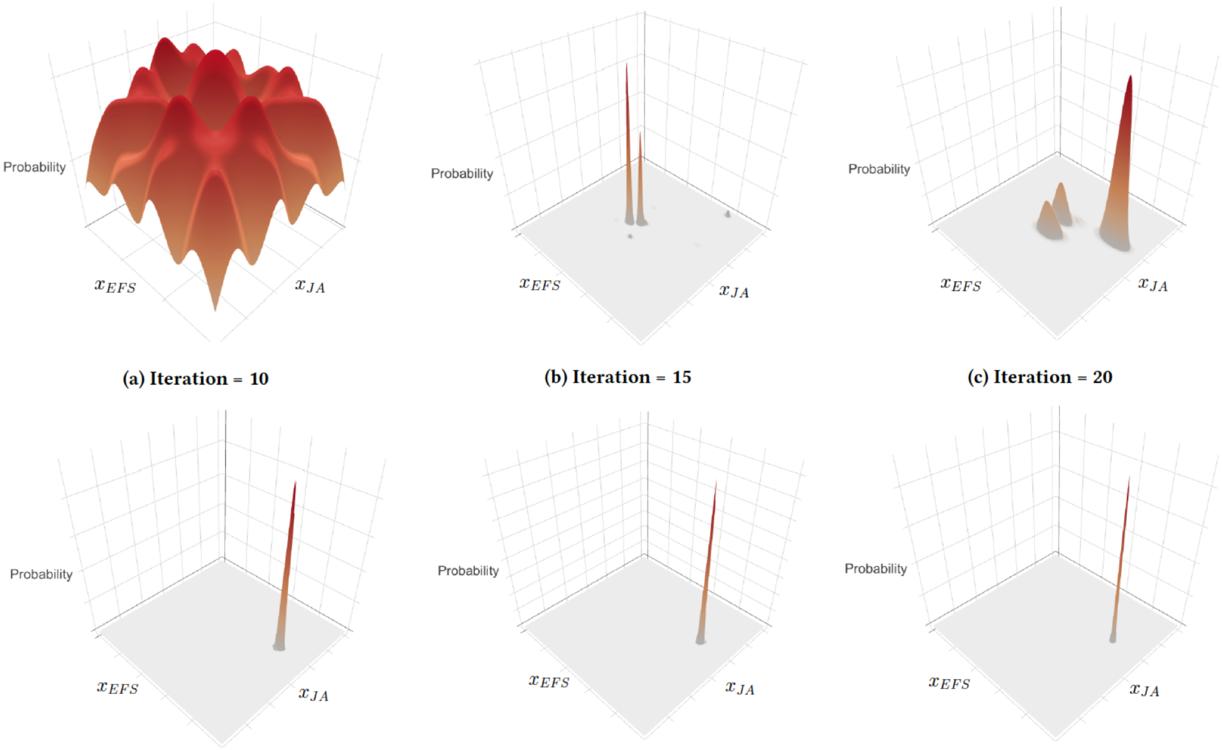
Online Architecture



Other Practical Design Considerations

- Consistency in user experience.
  - Randomize at member level instead of session level.
- Offline flow frequency
  - Batch computation where we collect data for an hour and run the offline flow each hour to update the sampling distribution.
- Assume (f_{VA}, f_{EFS}, f_{JA}) to be Independent
  - Works well in our setup. Joint modeling might reduce variance.
- Choice of business constraint thresholds.
  - Chosen to allow for a 1% drop.

Online Convergence Result



(a) Iteration = 10 (b) Iteration = 15 (c) Iteration = 20

(d) Iteration = 25 (e) Iteration = 30 (f) Iteration = 35

Online A/B Test Results

Table 1: Online A/B results for Online Parameter Selection in LinkedIn Feed Ranking

Metric	Lift (%) vs Control $\boldsymbol{x}_{c_1}$	Lift (%) vs Control $\boldsymbol{x}_{c_2}$
Viral Actions	+3.3%	+1.2%
Engaged Feed Sessions	-0.8%	0%
Job Applies	+12.8%	+6.4%

Key Takeaways

- Removes the human in the loop: Fully automatic process to find the optimal parameters.

- Drastically improves developer productivity.

- Can scale to multiple competing metrics.

- Future Direction
  - Create a dependent structure on different utilities to better model the variance.
  - Automatically identify the primary metric by understanding the models better.
  - Allow for relaxing the constraint to get large gain in primary metric.

LinkedIn