Online Parameter Selection for Web-based Ranking

DEEPAK AGARWAL, KINJAL BASU, SOUVIK GHOSH, YING XUAN, YANG YANG, LIANG ZHANG
LinkedIn Corporation

**LinkedIn Feed**

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

The Feed is the personalized home page of LinkedIn and contains a heterogeneous list of updates:

- Shares from a member’s connections.
- Recommendations including jobs, articles, connections, courses.
- Sponsored Content or Ads.

**Ranking Problem**

The ranking problem on the feed tries to balance three important metrics, Viral Actions (VA), Job Applies (JA), and Engaged Feed Session (EFS). For a member \( m \) the updates \( u \) in the feed is ranked according to

\[
S(m, u) = P_{VA}(m, u) + x_{EFS}P_{EFS}(m, u) + x_{JA}P_{JA}(m, u)
\]

(1)

The weight vector \( x = (x_{EFS}, x_{JA}) \) controls the balance of the metrics EFS, VA and JA. The business strategy is

\[
\max_x VA(x) \quad \text{s.t.} \quad EFS(x) \geq c_{EFS}, JA(x) \geq c_{JA}
\]

(2)

**Reformulation for Bayesian Optimization**

The optimal value of \( x \) (tuning parameters) changes over time. Example of changes can include new content types or updated relevance models. With every change engineers would manually find the optimal \( x \) by running multiple A/B tests and it is not the best use of engineering time.

- Let \( Y_{ij}^k(x) \in \{0, 1\} \) 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 JA.
- \( Y_{ij}^k(x) \sim Bin(n_i(x), \sigma(f_k(x))) \)

Based on this modeling we reformulate the original problem as

\[
\max_x \sigma(f_{VA}(x)) + \lambda(\sigma_{EFS}(x) - c_{EFS}) + \sigma_{JA}(x) - c_{JA})
\]

(3)

**Proposed Solution:** We solve the problem through an \( \epsilon \)-greedy Thompson Sampling Algorithm. Each function \( f_k \) is modelled as a Gaussian Process. We start with a random distribution on \( x \) and using the observed data, we estimate the posterior of each \( f_k \). We sample from the posterior and estimate the new distribution of the maximum \( x^* \). We continue this process till convergence.

**System Architecture**

**Simulation**

(a) Trimodal Shekel Function
(b) Decay of log relative square error

**Online Results**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Lift in control 1</th>
<th>Lift in control 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral Action</td>
<td>+3.3%</td>
<td>+1.2%</td>
</tr>
<tr>
<td>Engaged Feed Session</td>
<td>-0.8%</td>
<td>0 %</td>
</tr>
<tr>
<td>Job Applies</td>
<td>+12.8%</td>
<td>+6.4%</td>
</tr>
</tbody>
</table>