A Bandit Approach to Personalized News Article Recommendation

Qingyun Wu
News Recommendation Cycle
A K-armed Bandit Formulation

- A gambler must decide which of the K non-identical slot machines (we called them arms) to play in a sequence of trails in order to maximize total reward.

News Website ←→ gambler
Candidate news articles ←→ arms
User Click ←→ Reward

How to pull arms to maximize reward?

How to select articles to serve users to maximize user clicks
Ideal Solution

Pick \( \arg \max_a \mu_a \)

But we DO NOT know the mean.

Let’s estimate it

<table>
<thead>
<tr>
<th>Choices</th>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( X_5 )</th>
<th>( X_6 )</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>( a_2 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a_k )</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Time
Exploitation VS. Exploration

**Exploitation:** pull an arm for which we currently have the highest estimate of mean of reward

**Exploration:** Pull an arm we never pulled before

Not only look at the mean, but also the confidence!

\[
\text{Pick } \ arg \ max_a (\mu_a + \alpha \times UCB)
\]
UCB1

\[
\arg \max_a (\hat{\mu}_a + \sqrt{\frac{2 \ln T}{n_a}})
\]

LinUCB (Contextual)

- **Article feature**: URL categories, topic categories
- **User feature**: demographic information, geographic features, behavioral categories
LinUCB(Contextual)

Assumption

\[ E(y_{t,n} | x_{t,n}) = x_{t,n}^T \theta_n \]

Article Feature Vector

User preference

Parameter Estimation

\[ \hat{\theta}_n = A^{-1}b \]

\[ A_n = \lambda I + \sum_{t} x_{t,n} x_{t,n}^T \]

\[ b_n = \sum_{t} y_{t,n} x_{t,n} \]

Pick

\[ \arg\max_{a} (x_{t,n}^T \hat{\theta}_n + \alpha \sqrt{x_{t,n}^T (D_n^T D_n + I_d) x_{t,a}}) \]
From LinUCB to Collaborative-LinUCB

If user $i$ and user $j$ are connected by an edge, $W_{ij} > 0$

Otherwise $W_{ij} = 0$

Assumption

$$\sum_{i=1}^{N} W_{ij} = 1 \quad \sum_{j=1}^{N} W_{ij} = 1$$

$$E(r_{t,n} \mid x_{t,n}) = x_{t,n}^T \theta_n$$

$$\rightarrow$$

$$E(r_{t,n} \mid x_{t,n}) = x_{t,n}^T \sum_{j}^{N} W_{nj} \theta_j$$
Collaborative-LinUCB

Parameter Estimation

\[ \hat{\theta}_n = A_n^{-1} b_n \]

\[ A_n = \lambda I + \sum_{m=1}^{N} W_{mn}^2 \sum_t x_{tm} x_{tm}^T \]

\[ b_n = \sum_{m=1}^{N} W_{mn} \sum_t (x_{tm} y_{tm} - x_{tm} x_{tm}^T \sum_{j\neq n}^{N} W_{mj} \theta_j^U) \]

Make a choice

\[ \arg\max_a (x_{tn}^T \sum_{j=1}^{N} \hat{\theta}_{nj} + \alpha \sqrt{x_{tn}^T \sum_{j=1}^{N} W_{nj} A_j^{-1} x_{tn}}) \]
Performance Evaluation

Measurement criteria

Regret

\[ R_A(T) = E[\sum_{t} r_{t,a_t^*}] - E[\sum_{t} r_{t,a_t}] \]
Performance Evaluation

![Graphs showing performance evaluation results. The top graph displays regret vs iteration for CoLinUCB and LinUCB, with a noise scale of 0.01. The bottom graph illustrates square root L2 difference vs iteration for CoLinUCB_CoTheta, LinUCB_CoTheta, and CoLinUCB_Theta.]
Performance Evaluation

- **Noise scale = 0.1**
  - Regret
    - CoLinUCB
    - LinUCB

- **SqRoot L2 Diff**
  - CoLinUCB_CoTheta
  - LinUCB_CoTheta
  - CoLinUCB_Theta

Iteration vs. Regret and SqRoot L2 Diff for different methods with noise scale 0.1.
Performance Evaluation

Noise scale = 1

Regret vs Iteration

SqRoot L2 Diff vs Iteration
Summary

UCB1

LinUCB

CoLinUCB