A Contextual-Bandit Approach to Personalized News Article Recommendation

Lihong Li, Wei Chu, John Langford, Robert E. Schapire

Presenter: 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 $\leftarrow$ gambler
Candidate news articles $\leftarrow$ arms
User Click $\leftarrow$ Reward

How to pull arms to maximize reward?
A K-armed Bandit formulation

Setting

- Set of K choices (arms)
- Each choice $a$ is associated with an unknown probability distribution $p_a$ supported in $[0,1]$
- Play the game for $T$ rounds

- In each round $t$
  1. We pick article $j$
  2. We observe random sample $X_t$ from $p_j$

Our Goal: maximize $\sum_{t=1}^{T} X_t$
Ideal Solution

\[ \text{Pick } \arg \max_a \mu_a \]

But we DO NOT know the mean.
Every time we pull an arm we learn a bit more about the distribution.
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

Extreme examples:

**Greedy Strategy:** Take the arm with the highest average reward

**Random Strategy:** Randomly choose an arm

Too confident  Too unconfident
How to make trade off

Don’t just look at the mean (that’s the expected reward), but also the confidence!
UCB (Upper Confidence Bound) algorithm

Pick \( \arg \max_a (\hat{\mu}_a + \alpha \cdot \text{Varance}) \)

\[ \downarrow \]

Pick \( \arg \max_a (\hat{\mu}_a + \alpha \cdot UCB) \)

Confidence Interval is a range of values within which we are sure the mean lies with a certain probability.

UCB1 \( \arg \max_a (\hat{\mu}_a + \sqrt{\frac{2 \ln T}{n_a}}) \)

Reference: Finite-Analysis of the Multi-armed Bandit Problem, Peter Auer, Nicolo Cesa-Bianchi, Paul Fischer
http://homes.di.unimi.it/~cesabian/Pubblicazioni/ml-02.pdf
Make use of Contextual Information

- **User feature**: demographic information, geographic features, behavioral categories

- **Article feature**: URL categories, topic categories

**Assumption about the reward:**

The expected reward of an arm \( a \) is linear in its \( d \)-dimensional feature \( x_{t,a} \), with some unknown coefficient vector \( \theta^*_a \), namely, for all \( t \),

\[
E(r_{t,a} \mid x_{t,a}) = x_{t,a}^T \theta^*_a
\]
UCB(Upper Confidence Bound) algorithm

**Assumption**
\[ E(r_{t,a} \mid x_{t,a}) = x_{t,a}^T \theta_a^* \]

**Parameter Estimation**
\[ \hat{\theta}_a = (D_a^T D_a + I_d)^{-1} D_a^T c_a \]  
(Ridge Regression)

**Bound of the variance**
\[ x_{t,a}^T \hat{\theta}_a - E(r_{t,a} \mid x_{t,a}) \leq \alpha \sqrt{x_{t,a}^T (D_a^T D_a + I_d)^{-1} x_{t,a}} \]

Bound we need!!!

**Pick**
\[ \arg\max_a (x_{t,a}^T \hat{\theta}_a + \alpha \sqrt{x_{t,a}^T (D_a^T D_a + I_d) x_{t,a}}) \]
### Performance Evaluation

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>size = 100%</th>
<th></th>
<th>size = 30%</th>
<th></th>
<th>size = 20%</th>
<th></th>
<th>size = 10%</th>
<th></th>
<th>size = 5%</th>
<th></th>
<th>size = 1%</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>deploy</td>
<td>learn</td>
<td>deploy</td>
<td>learn</td>
<td>deploy</td>
<td>learn</td>
<td>deploy</td>
<td>learn</td>
<td>deploy</td>
<td>learn</td>
<td>deploy</td>
<td>learn</td>
</tr>
<tr>
<td>$\epsilon$-greedy</td>
<td>1.596</td>
<td>1.326</td>
<td>1.541</td>
<td>1.326</td>
<td>1.549</td>
<td>1.273</td>
<td>1.465</td>
<td>1.326</td>
<td>1.409</td>
<td>1.292</td>
<td>1.234</td>
<td>1.139</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>UCB</td>
<td>1.594</td>
<td>1.569</td>
<td>1.582</td>
<td>1.535</td>
<td>1.569</td>
<td>1.488</td>
<td>1.541</td>
<td>1.446</td>
<td>1.541</td>
<td>1.465</td>
<td>1.354</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>18.3%</td>
<td>2.7%</td>
<td>15.8%</td>
<td>1.3%</td>
<td>16.9%</td>
<td>5.2%</td>
<td>9%</td>
<td>9.4%</td>
<td>13.4%</td>
<td>9.7%</td>
<td>7.1%</td>
</tr>
<tr>
<td>$\epsilon$-greedy (seg)</td>
<td>1.742</td>
<td>1.446</td>
<td>1.652</td>
<td>1.46</td>
<td>1.585</td>
<td>1.119</td>
<td>1.474</td>
<td>1.284</td>
<td>1.407</td>
<td>1.281</td>
<td>1.245</td>
<td>1.072</td>
</tr>
<tr>
<td></td>
<td>9.1%</td>
<td>9%</td>
<td>7.2%</td>
<td>10.1%</td>
<td>2.3%</td>
<td>-12%</td>
<td>0.6%</td>
<td>-3.1%</td>
<td>0%</td>
<td>-0.8%</td>
<td>0.9%</td>
<td>-5.8%</td>
</tr>
<tr>
<td>UCB (seg)</td>
<td>1.781</td>
<td>1.677</td>
<td>1.742</td>
<td>1.555</td>
<td>1.689</td>
<td>1.446</td>
<td>1.636</td>
<td>1.529</td>
<td>1.532</td>
<td>1.32</td>
<td>1.398</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>11.6%</td>
<td>26.5%</td>
<td>13%</td>
<td>17.3%</td>
<td>9%</td>
<td>13.6%</td>
<td>11.7%</td>
<td>15.3%</td>
<td>8.7%</td>
<td>2.2%</td>
<td>13.3%</td>
<td>9.7%</td>
</tr>
<tr>
<td>$\epsilon$-greedy (disjoint)</td>
<td>1.769</td>
<td>1.309</td>
<td>1.686</td>
<td>1.337</td>
<td>1.624</td>
<td>1.529</td>
<td>1.529</td>
<td>1.451</td>
<td>1.432</td>
<td>1.345</td>
<td>1.262</td>
<td>1.183</td>
</tr>
<tr>
<td></td>
<td>10.8%</td>
<td>-1.2%</td>
<td>9.4%</td>
<td>0.8%</td>
<td>4.8%</td>
<td>20.1%</td>
<td>4.4%</td>
<td>9.4%</td>
<td>1.6%</td>
<td>4.1%</td>
<td>2.3%</td>
<td>3.9%</td>
</tr>
<tr>
<td>LINUCB (disjoint)</td>
<td>1.795</td>
<td>1.647</td>
<td>1.719</td>
<td>1.507</td>
<td>1.714</td>
<td>1.384</td>
<td>1.655</td>
<td>1.387</td>
<td>1.574</td>
<td>1.245</td>
<td>1.382</td>
<td>1.197</td>
</tr>
<tr>
<td></td>
<td>12.5%</td>
<td>24.2%</td>
<td>11.6%</td>
<td>13.7%</td>
<td>10.7%</td>
<td>8.7%</td>
<td>13%</td>
<td>4.6%</td>
<td>11.7%</td>
<td>-3.5%</td>
<td>12%</td>
<td>5.1%</td>
</tr>
<tr>
<td>$\epsilon$-greedy (hybrid)</td>
<td>1.739</td>
<td>1.521</td>
<td>1.68</td>
<td>1.345</td>
<td>1.636</td>
<td>1.449</td>
<td>1.58</td>
<td>1.348</td>
<td>1.465</td>
<td>1.415</td>
<td>1.342</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>9%</td>
<td>14.7%</td>
<td>9%</td>
<td>1.4%</td>
<td>5.6%</td>
<td>13.8%</td>
<td>7.8%</td>
<td>1.7%</td>
<td>4%</td>
<td>9.5%</td>
<td>8.8%</td>
<td>5.4%</td>
</tr>
<tr>
<td>LINUCB (hybrid)</td>
<td>1.733</td>
<td>1.663</td>
<td>1.691</td>
<td>1.591</td>
<td>1.708</td>
<td>1.619</td>
<td>1.675</td>
<td>1.535</td>
<td>1.588</td>
<td>1.507</td>
<td>1.482</td>
<td>1.446</td>
</tr>
<tr>
<td></td>
<td>8.4%</td>
<td>25.4%</td>
<td>9.7%</td>
<td>20%</td>
<td>10.3%</td>
<td>27.2%</td>
<td>14.3%</td>
<td>15.8%</td>
<td>12.7%</td>
<td>16.6%</td>
<td>20.1%</td>
<td>27%</td>
</tr>
</tbody>
</table>

Table 1: Performance evaluation: CTRs of all algorithms on the one-week evaluation dataset in the deployment and learning bucket (denoted by “deploy” and “learn” in the table, respectively). The numbers with a percentage is the CTR lift compared to $\epsilon$-greedy.
Summary

- Model news recommendation as a K-armed Bandit Problem
- UCB-type Algorithm
- Take Contextual Information in to consideration