Bandit Models for Recommendations

Maryam Tavakol

Computer Science 8, Artificial Intelligence

July 2019
Recommender Systems
Recommender Systems

• e-commerce
Recommender Systems

- e-commerce
- entertainment
Recommender Systems

- e-commerce
- entertainment
- social networks
Recommender Systems

- e-commerce
- entertainment
- social networks
- online content
Multi-armed Bandits
Multi-armed Bandits

- Which one to play?
Multi-armed Bandits

• Which one to play?
Multi-armed Bandits

• Which one to play?
Multi-armed Bandits

• Which one to play?

10$, 0$, -5$, 1$
-8$, 4$, 0$, 1$

• strategy: to better balance exploration vs. exploitation
Bandits in Recommendation
Bandits in Recommendation

• A user views the following item:
Bandits in Recommendation

- A user views the following item:

- Now what to recommend?
Bandits in Recommendation

- A user views the following item:

- Now what to recommend?
Bandits in Recommendation

• A user views the following item:

• Now what to recommend?

what about the context?

<table>
<thead>
<tr>
<th>Context</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Male, 50, ...)</td>
<td>1.0</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>(Female, 18, ...)</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>(Female, 48, ...)</td>
<td>0.5</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Contextual Bandits

• In contextual bandit, the system receives an additional information (=context) at every step

  • the reward is represented as a function of context

• sequential procedure:

  • repeat:
    1. learner presented with context
    2. learner chooses an action
    3. learner observes reward (but only for chosen action)

• goal: learn to choose actions to maximize rewards
LinUCB for News Recommendation
LinUCB for News Recommendation

- The payoff function is defined by a linear model

\[ \mathbb{E}[r_{t,a} | x_t] = x_t^\top w \]
LinUCB for News Recommendation

• The payoff function is defined by a linear model

\[ \mathbb{E}[r_{t,a} | \mathbf{x}_t] = \mathbf{x}_t^\top \mathbf{w} \]

to be learned
LinUCB for News Recommendation

• The payoff function is defined by a linear model

\[ \mathbb{E}[r_{t,a} | x_t] = x_t^\top w \]

to be learned

• arm selection strategy:

• the arm with the highest mean reward + confidence interval
LinUCB for News Recommendation

• The payoff function is defined by a linear model

\[ \mathbb{E}[r_{t,a} | x_t] = x_t^T w \]

• arm selection strategy:
  • the arm with the highest mean reward + confidence interval

exploitation exploration
LinUCB for News Recommendation

• The payoff function is defined by a linear model

\[ E[r_{t,a} \mid x_t] = x_t^\top w \]

• arm selection strategy:
  • the arm with the highest mean reward + confidence interval

Upper Confidence Bound (UCB)
Thanks for Your Attention

• Read more…

• contextual bandits in recommendation:
  


• off-line evaluation of recommendations using bandits:


Questions?