Whole-Chain Recommendations

Xiangyu Zhao  Long Xia  Lixin Zou  Hui Liu  Dawei Yin  Jiliang Tang

1: Michigan State University  2: York University  3: Baidu Inc
Background

- Users sequentially interact with multiple scenarios
  - Each scenario has different objective
Motivation

- Optimizing each recommender agent for each scenario
  - Ignoring sequential dependency
  - Missing information
  - Sub-optimal overall objective
Whole-Chain Recommendation

- **Goal**
  - Jointly optimizing multiple recommendation strategies
  - Maximizing the overall performance of the whole session

- **Advantages**
  - Agents are sequentially activated
  - Agents share the same memory
  - Agents work collaboratively

- **Actor-Critic**
  - Actor: recommender agent in one scenario
  - Critic: controlling actors
Individual Actor

- **Goal**
  1. Capturing users’ preference from their browsing history (state)
  2. Generating recommendations (action)

- **Encoder-Decoder**
Global Critic (Q-function)

- **Goal**
  - Controlling all actors to work collaboratively → optimize global performance

- **Challenge**
  - How to capture user’s attention pattern in different scenarios?

- **Solution**
  - Separate attention mechanisms
Data Science and Engineering Lab

Optimization

Entrance Page

Item Detail Page

Actor

skip

click

Entrance Page

§ 1st row: skip behavior
2nd row: click behavior
3rd row: leave behavior

\[ y_t = \begin{bmatrix} p_m(s_t, a_t) \cdot Q_m(s_{t+1}, \pi_m(s_{t+1})) + p_c(s_t, a_t) \cdot (r_t + \gamma Q_c(s_{t+1}, \pi_c(s_{t+1})) \cdot \gamma Q_m(s_{t+1}, \pi_m(s_{t+1})) \cdot r_t) m \end{bmatrix} \]
Why Model-based RL?

- Advantages
  - Reducing training data amount requirement
  - Performing accurate optimization of the Q-function

\[
y_t = \left[ p^s_m(s_t, a_t) \cdot \gamma Q_{\mu'}(s_{t+1}, \pi'_m(s_{t+1})) \right.
+ p^c_m(s_t, a_t) \cdot (r_t + \gamma Q_{\mu'}(s_{t+1}, \pi'_d(s_{t+1})))
+ p^l_m(s_t, a_t) \cdot r_t \right] \mathbf{1}_m
+ \left[ p^c_d(s_t, a_t) \cdot (r_t + \gamma Q_{\mu'}(s_{t+1}, \pi'_d(s_{t+1}))) \right.
+ p^s_d(s_t, a_t) \cdot \gamma Q_{\mu'}(s_{t+1}, \pi'_m(s_{t+1}))
+ p^l_d(s_t, a_t) \cdot r_t \right] \mathbf{1}_d
\]
## Experiment on JD.com Data

### Baselines
- Wide&Deep
- DeepFM
- GRU4Rec
- DDPG
- MA-RDPG

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Metrics</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>W&amp;D</td>
</tr>
<tr>
<td>Entrance Page</td>
<td>MAP</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>improv.(%)</td>
<td>18.87</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.000</td>
</tr>
<tr>
<td>Entrance Page</td>
<td>NDCG@40</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>improv.(%)</td>
<td>19.05</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.000</td>
</tr>
<tr>
<td>Item Detail Page</td>
<td>MAP</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>improv.(%)</td>
<td>18.52</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.000</td>
</tr>
<tr>
<td>Item Detail Page</td>
<td>NDCG@40</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>improv.(%)</td>
<td>18.67</td>
</tr>
<tr>
<td></td>
<td>p-value</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Experiment on Simulated Online Environment

- **Baselines**
  - Wide&Deep
  - DeepFM
  - GRU4Rec
  - DDPG
  - MA-RDPG

- **Variants**
  - DC-o: one-agent
  - DC-f: model-free
Thanks

zhaoxi35@msu.edu