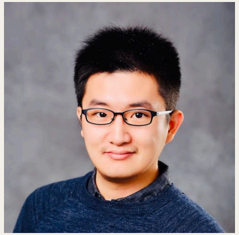
Whole-Chain Recommendations



Xiangyu Zhao

Long Xia

Lixin Zou

Hui Liu

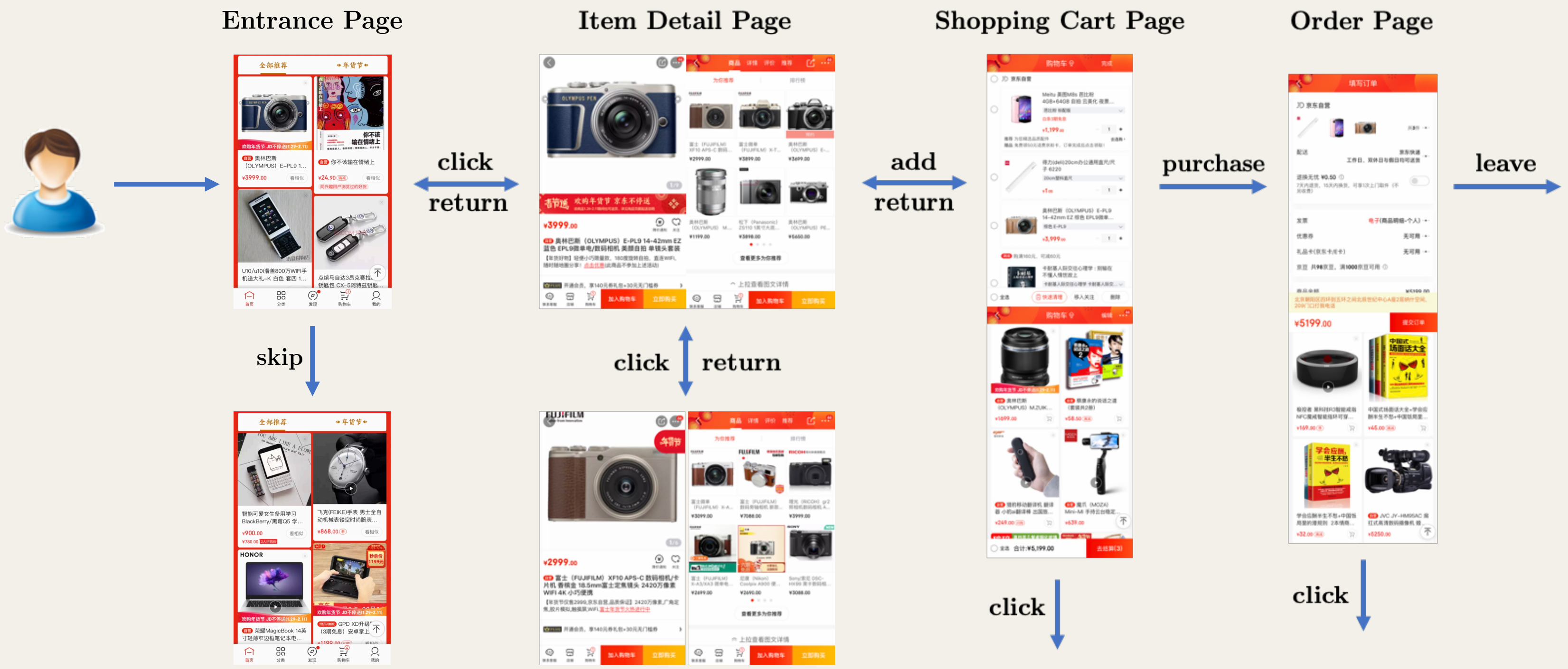
Dawei Yin

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

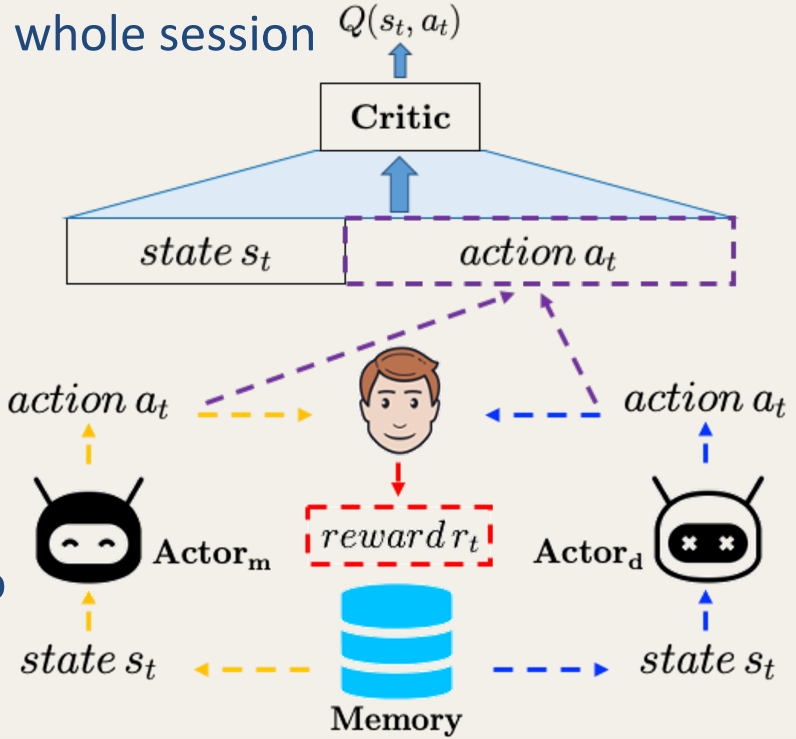
- Optimizing each recommender agent for each scenario
  - Ingoring sequential dependency
  - Missing information
  - Sub-optimal overall objective

Entrance Page

Item Detail Page

Whole-Chain Recommendation

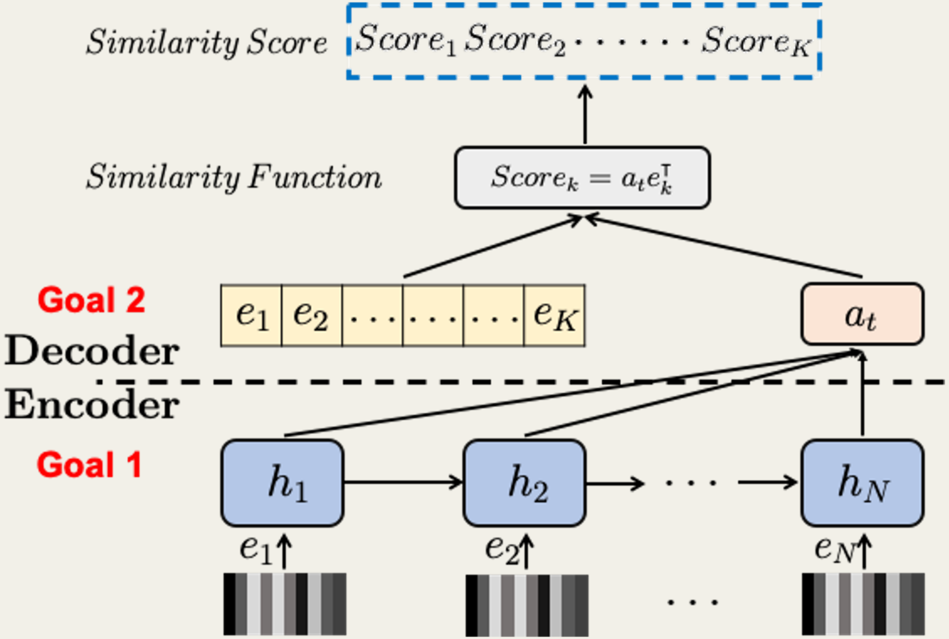
- Goal
  - Jointly optimizing multiple recommendation strategies
  - Maximizing the overall performance of the whole session
- Advantages
  - Agents are sequentially activated
  - Agents share the same memory
  - Agents work collaboratively
- Actor-Critic
  - Actor: recommender agent in one scenario
  - Critic: controlling actors



Individual Actor

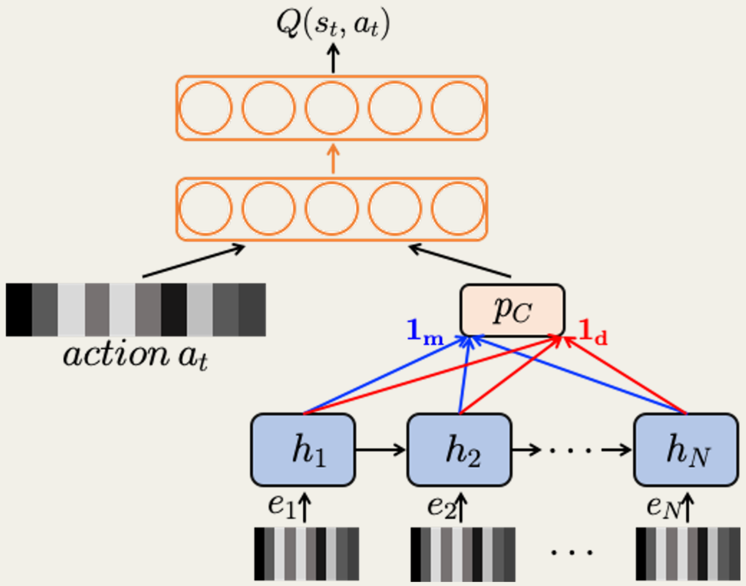
- Goal
  1. Capturing users' preference from their browsing history (state)
  2. Generating recommendations (action)

- Encoder-Decoder

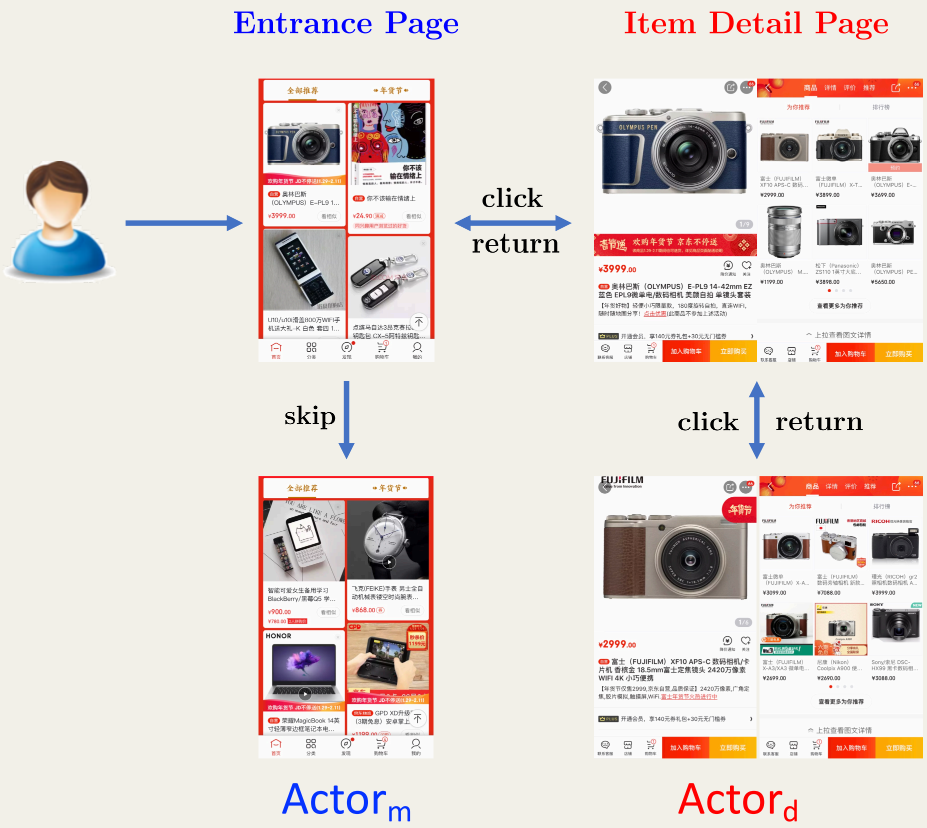


Global Critic (Q-function)

- 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



Optimization



Entrance Page

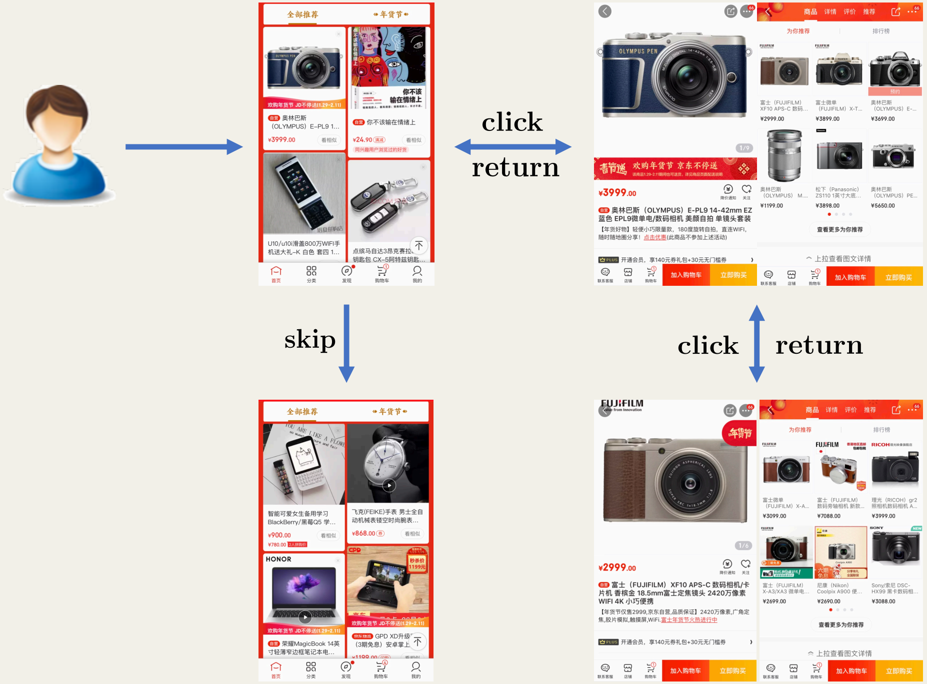
$$\begin{aligned} y_t = \Big[& p_m^s(s_t, a_t) \cdot \gamma Q_{\mu'}\big(s_{t+1}, \pi'_m(s_{t+1})\big) \\ + \ & p_m^c(s_t, a_t) \cdot \Big(r_t + \gamma Q_{\mu'}\big(s_{t+1}, \pi'_d(s_{t+1})\big)\Big) \\ + \ & p_m^l(s_t, a_t) \cdot r_t \Big] \mathbf{1}_{\mathbf{m}} \end{aligned}$$

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

Optimization

Entrance Page

Item Detail Page



Actor$_m$

Actor$_d$

Entrance Page

$$
\begin{aligned}
y_t = \Big[& p_m^s(s_t, a_t) \cdot \gamma Q_{\mu'}(s_{t+1}, \pi'_m(s_{t+1})) \\
& + p_m^c(s_t, a_t) \cdot \big(r_t + \gamma Q_{\mu'}(s_{t+1}, \pi'_d(s_{t+1}))\big) \\
& + p_m^l(s_t, a_t) \cdot r_t \Big] \mathbf{1}_{\mathrm{m}} \\
& + \Big[p_d^c(s_t, a_t) \cdot \big(r_t + \gamma Q_{\mu'}(s_{t+1}, \pi'_d(s_{t+1}))\big) \\
& + p_d^s(s_t, a_t) \cdot \gamma Q_{\mu'}(s_{t+1}, \pi'_m(s_{t+1})) \\
& + p_d^l(s_t, a_t) \cdot r_t \Big] \mathbf{1}_{\mathrm{d}}
\end{aligned}
$$

Item Detail Page

Why Model-based RL?

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

$$
\begin{aligned}
y_t = &\Big[p_m^s(s_t, a_t) \cdot \gamma Q_{\mu'}(s_{t+1}, \pi'_m(s_{t+1})) \\
&+ \; p_m^c(s_t, a_t) \cdot \big(r_t + \gamma Q_{\mu'}(s_{t+1}, \pi'_d(s_{t+1}))\big) \\
&+ \; p_m^l(s_t, a_t) \cdot r_t\Big]\mathbf{1_m} \\
&+ \Big[p_d^c(s_t, a_t) \cdot \big(r_t + \gamma Q_{\mu'}(s_{t+1}, \pi'_d(s_{t+1}))\big) \\
&+ \; p_d^s(s_t, a_t) \cdot \gamma Q_{\mu'}(s_{t+1}, \pi'_m(s_{t+1})) \\
&+ \; p_d^l(s_t, a_t) \cdot r_t\Big]\mathbf{1_d}
\end{aligned}
$$

Model-based

Experiment on JD.com Data

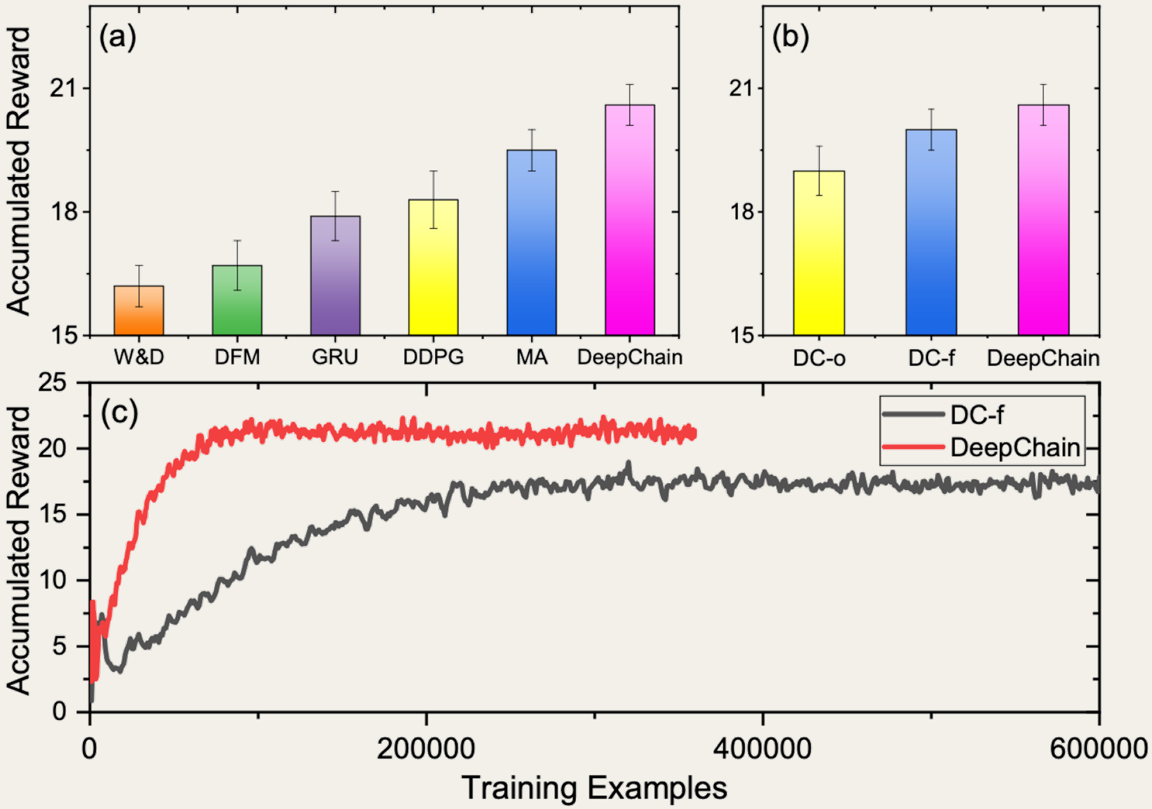
- Baselines
  - Wide&Deep
  - DeepFM
  - GRU4Rec
  - DDPG
  - MA-RDPG

Scenarios	Metrics	Algorithms					
		W&D	DFM	GRU	DDPG	MA	DeepChain
Entrance Page	MAP	0.106	0.108	0.113	0.117	0.121	**0.126**
	improv.(%)	18.87	16.67	11.50	7.693	4.132	-
	p-value	0.000	0.000	0.000	0.000	0.003	-
Entrance Page	NDCG@40	0.189	0.193	0.201	0.209	0.215	**0.225**
	improv.(%)	19.05	16.58	11.95	7.656	4.651	-
	p-value	0.000	0.000	0.000	0.000	0.003	-
Item Detail Page	MAP	0.081	0.083	0.086	0.090	0.093	**0.096**
	improv.(%)	18.52	15.66	11.63	6.667	3.226	-
	p-value	0.000	0.000	0.000	0.000	0.006	-
Item Detail Page	NDCG@40	0.166	0.169	0.176	0.183	0.190	**0.197**
	improv.(%)	18.67	16.57	11.93	7.650	3.684	-
	p-value	0.000	0.000	0.000	0.000	0.005	-

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