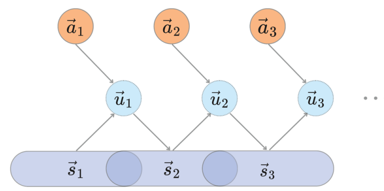
Representation Learning in Low-rank Slate-based Recommender System

Yijia Dai, Wen Sun
Department of Computer Science, Cornell University
yd73@cornell.edu

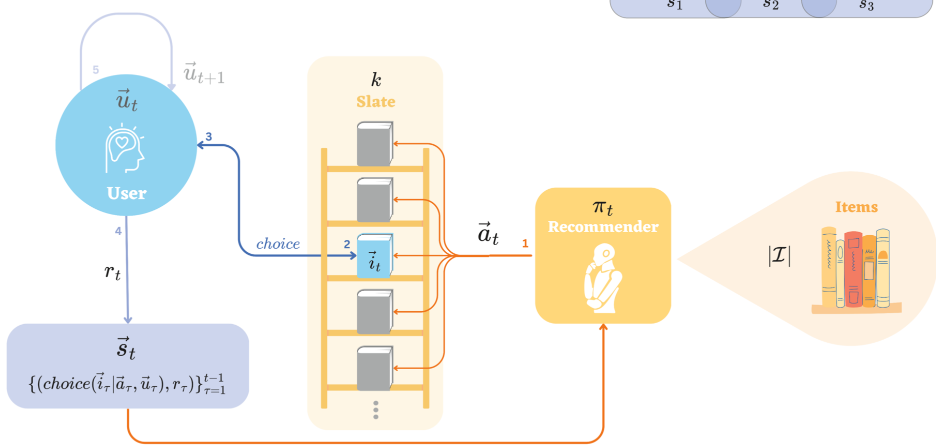
Using RL methods in recommender systems faces an issue regarding the large observation and action space, and doing efficient exploration becomes a harder question. Prior RL methods in recommender systems often overlook exploration or use ε-greedy and Boltzmann exploration.

Low-rank Representation

$$\forall \vec{s}, \vec{s'} \in \mathcal{S}, \vec{a} \in \mathcal{A} : P(\vec{s'}|\vec{s}, \vec{a}) = \mu(\vec{s'})^\top \phi(\vec{s}, \vec{a})$$



$\vec{u} = \phi(\vec{s}, \vec{a})$



Simulation is done using *Recsim*, with customized user choice model, user transition, observed states, sampling method and recommender algorithm.

From $|\mathcal{A}| = \binom{|\mathcal{I}|}{k}$ to $k|\mathcal{I}|$

- Assume the reward and transition depend only on the item that is consumed by the user on slate, i.e.,

$$r(\vec{s}, \vec{a}) = \sum_{\vec{i} \in \vec{a}} P(\vec{i}|\vec{s}, \vec{a}) r(\vec{s}, \vec{i})$$
$$P(\vec{s'}|\vec{s}, \vec{a}) = \sum_{\vec{i} \in \vec{a}} P(\vec{i}|\vec{s}, \vec{a}) P(\vec{s'}|\vec{s}, \vec{i})$$

- Can we shrink uniform action space from $\binom{|\mathcal{I}|}{k}$?

👀 User is indifferent to different slates, if there is a high chance to select the same item.

- Define $U(\mathcal{A})$ to be: 1. randomly pick an item $\vec{i}$; 2. assemble rest of the slate such that $\vec{i}$ has probability at least $\frac{1}{k}$ to be chosen by user.

- User select an item $\vec{i} \in \mathcal{I}$ with probability at least $\frac{1}{k|\mathcal{I}|}$. By pigeonhole principle, every $k|\mathcal{I}| + 1$ uniform actions lead to at least one duplicate action in this space.

Efficient Exploration

★ Rep-UCB-Rec has sample complexity of $O\left(\frac{d^4 k^2 |\mathcal{I}|^2}{\epsilon^2 (1-\gamma)^5}\right)$.

for episode $n = 1, \dots, N$ **do**

Collect a tuple $(\vec{s}, \vec{a}, \vec{s'}, \vec{a'}, \vec{\tilde{s}})$ with

$$\vec{s} \sim d_{P^\star}^{\pi_{n-1}}, \vec{a} \sim U(\mathcal{A}),$$
$$\vec{s'} \sim P^\star(\cdot|\vec{s}, \vec{a}), \vec{a'} \sim U(\mathcal{A}), \vec{\tilde{s}} \sim P^\star(\cdot|\vec{s}, \vec{a})$$

Update datasets

$$\mathcal{D}_n = \mathcal{D}_{n-1} + (\vec{s}, \vec{a}, \vec{s'}), \mathcal{D}'_n = \mathcal{D}'_{n-1} + (\vec{s'}, \vec{a'}, \vec{\tilde{s}})$$

Learn representation via ERM (i.e., MLE)

$$\hat{P}_n := (\hat{\mu}_n, \hat{\phi}_n)$$
$$= \arg \max_{(\mu, \phi) \in \mathcal{M}} E_{\mathcal{D}_n + \mathcal{D}'_n} \left[\ln \sum_{\vec{i} \in \vec{a}} \mu^\top(s') P(\vec{i}|\vec{s}, \vec{a}) \phi(\vec{s}, \vec{i}) \right]$$

Update empirical covariance matrix

$$\hat{\Sigma}_n = \sum_{\vec{s}, \vec{a} \in \mathcal{D}} \hat{\varphi}_n(\vec{s}, \vec{a}) \hat{\varphi}_n(\vec{s}, \vec{a})^\top + \lambda_n I$$
$$\text{where } \hat{\varphi}_n(\vec{s}, \vec{a}) := \sum_{\vec{i} \in a} P(\vec{i}|\vec{s}, \vec{a}) \hat{\phi}_n(\vec{s}, \vec{i})$$

Set the exploration bonus

$$\hat{b}_n(\vec{s}, \vec{a}) := \min\left(\alpha_n \sqrt{\hat{\varphi}_n(\vec{s}, \vec{a})^\top \hat{\Sigma}_n^{-1} \hat{\varphi}_n(\vec{s}, \vec{a})}, 2\right)$$

Update policy

$$\pi_n = \arg\max_\pi V^\pi_{\hat{P}_n, r + \hat{b}_n}$$