UNIVERSITY OF AMSTERDAM

Keeping Dataset Biases out of the Simulation

A Debiased Simulator for Reinforcement Learning based Recommender Systems

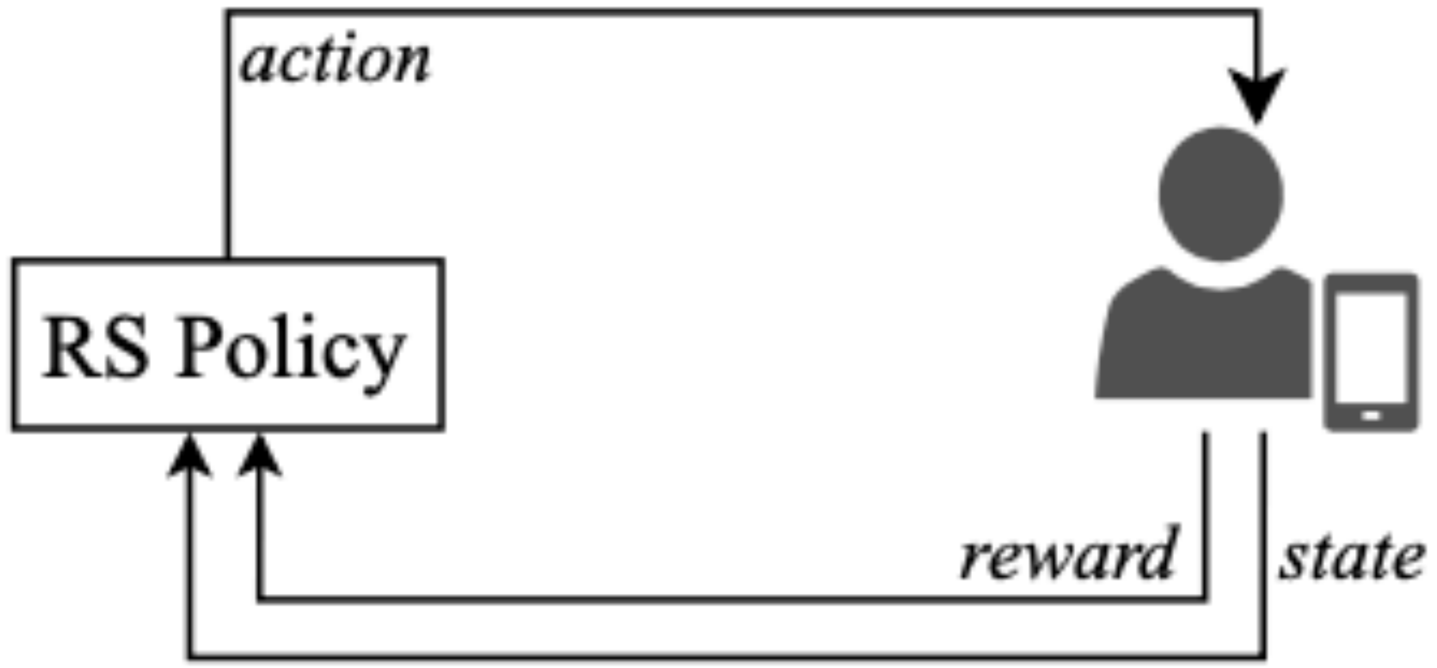
Jin Huang, Harrie Oosterhuis, Maarten de Rijke, Herke van Hoof

22 Sep, 2020

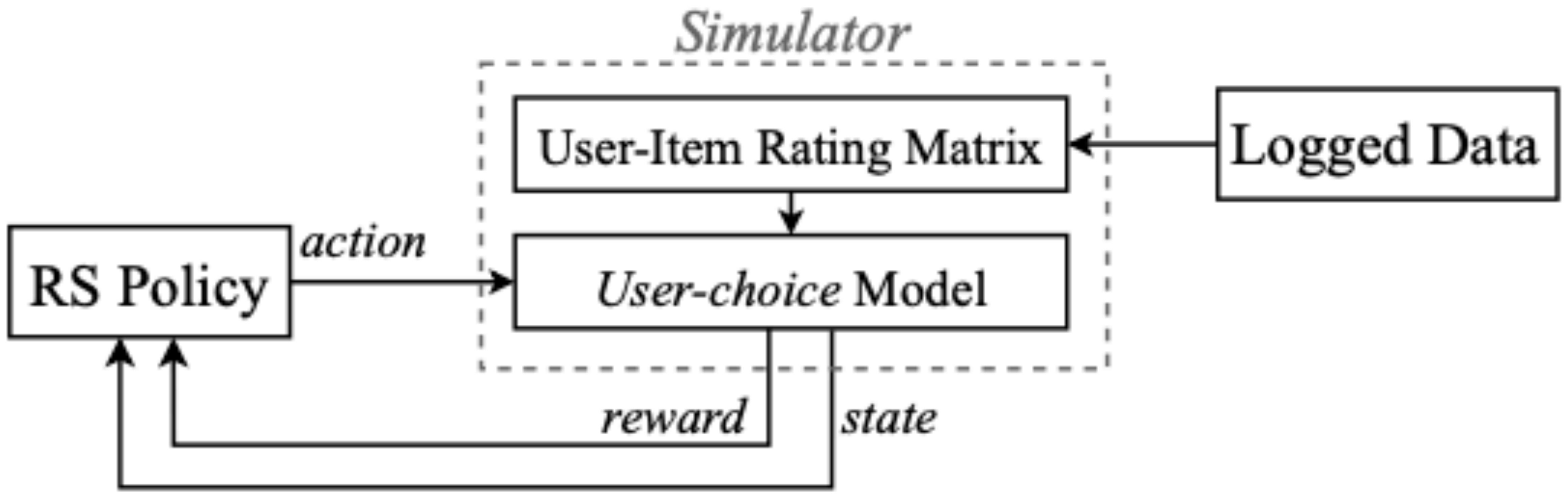
Main Contributions

- Simulator for RL-based RSs
  - Problem: the biases in logged data
  - Solution: Debiasing simulators to mitigate the effect of bias
- Simulation evaluation
  - A simulation evaluation method based on the performance of their produced policy to analyze the effect of bias on RL4Rec.
- A **S**imulator for **OF**fline le**A**rning and evaluation (SOFA), the first simulator for RL4Rec with correcting for bias in logged data.

Background: Reinforcement Learning for Recommendation **(RL4Rec)**



(a) RL4Rec online.

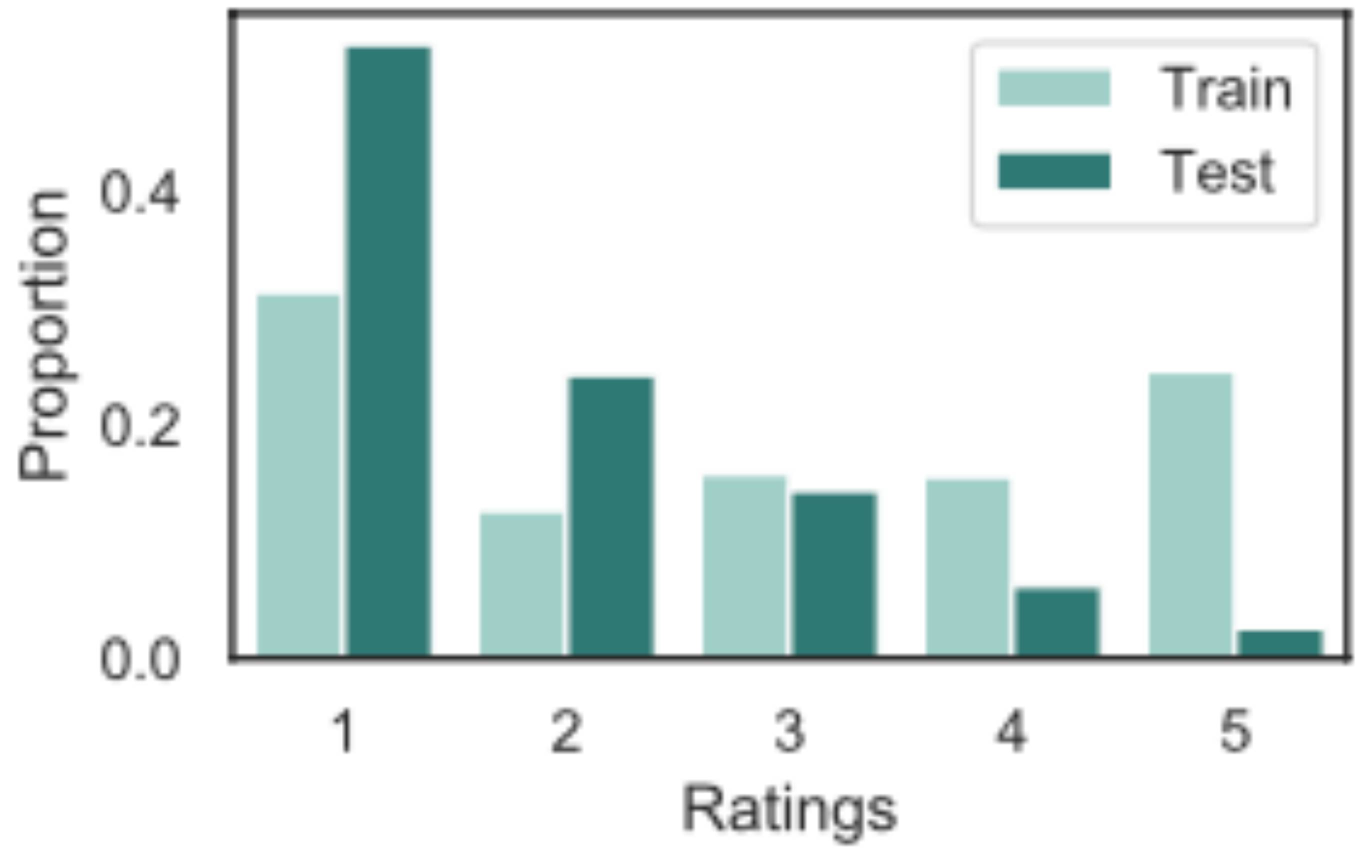


(b) RL4Rec with a simulator.

- Markov Decision Process (MDP)
  - State Space: {historical items, historical feedbacks}
  - Action Space: Item set *I*
  - *Reward: user feedback*
  - *Transition Probabilities*
  - *Discount Factor*

Background: Interaction Bias in Logged data

- Positivity bias in Yahoo!R3



- Missing Not At Random (MNAR) $P(y_{u,i}, o_{u,i}) = \underbrace{P(o_{u,i} \mid y_{u,i})}_{\text{Probability of observance}} \underbrace{P(y_{u,i})}_{\text{Probability of a rating}}$

(i) No Bias $\forall (u, u') \in U, (i, i') \in I, \left(P(o_{u,i}) = P(o_{u',i'})\right)$

(ii) Positivity Bias $\forall u \in U, (i, i') \in I, \left(y_{u,i} > y_{u,i'} \rightarrow P(o_{u,i}) > P(o_{u,i'})\right)$

- Effect of Bias
  - Example: estimate the average rating of an item $\texttt{avg}(i) = \frac{1}{N}\sum_{u \in U} y_{u,i}$

    The true average rating:

$$\mathbb{E}_o[\widehat{\texttt{avg}}(i)] = \frac{1}{\sum_{u \in U} P(o_{u,i} = 1)} \sum_{u \in U} P(o_{u,i} = 1) \cdot y_{u,i}$$

The naive (uncorrected) estimation is

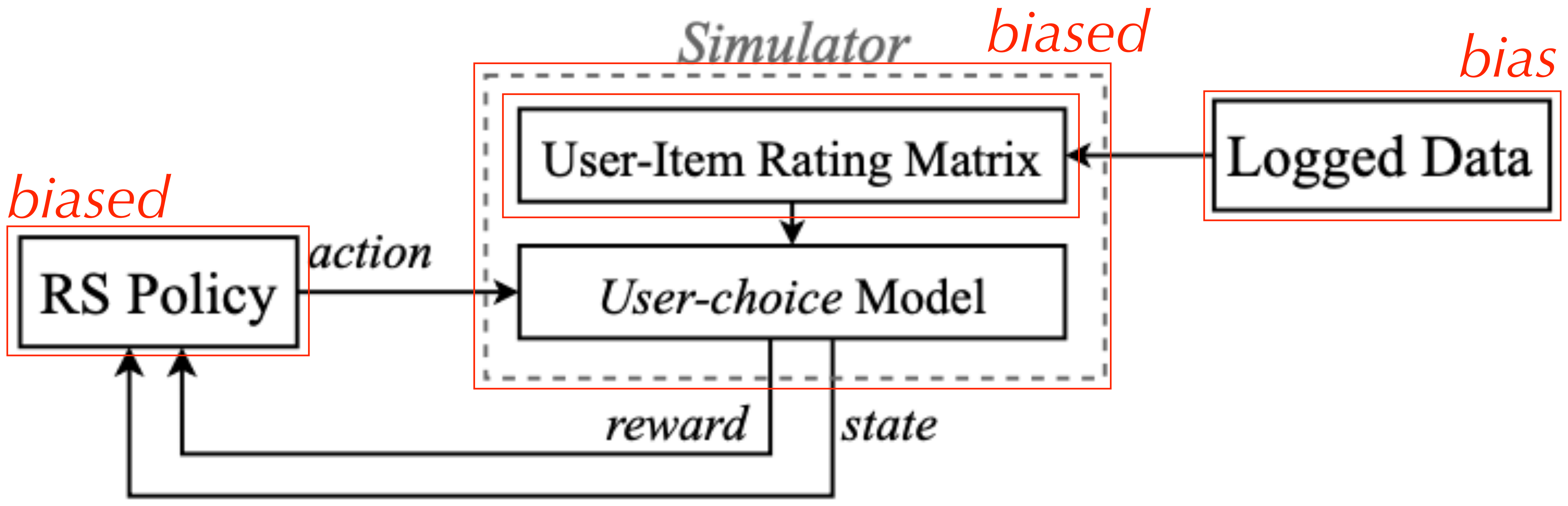
(i) No Bias $\forall (u, u') \in U, (i, i') \in I, \left(P(o_{u,i}) = P(o_{u',i'})\right)$ $\mathbb{E}_o[\widehat{\texttt{avg}}(i)] = \texttt{avg}(i)$

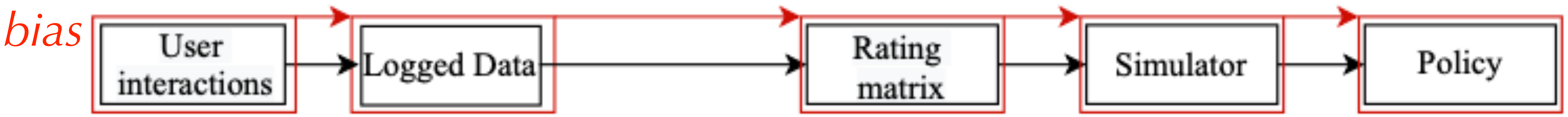
(iii) Positivity Bias $\forall u \in U, (i, i') \in I, \left(y_{u,i} > y_{u,i'} \rightarrow P(o_{u,i}) > P(o_{u,i'})\right)$ $\mathbb{E}_o[\widehat{\texttt{avg}}(i)] \geq \texttt{avg}(i)$

A Novel Method for Debiasing Simulators

Debiasing a Simulator

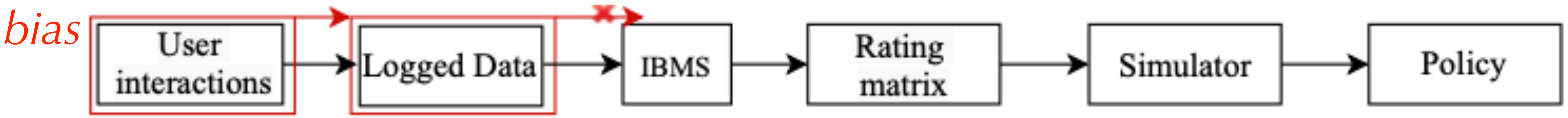
- RL4Rec simulators





Problem: simulated user behavior should not be affected by the way a dataset was logged.

- *Intermediate Bias Mitigation Step* (IBMS)



- Applied debiasing method in IBMS

Inverse Propensity Scoring (IPS)

- Standard rating prediction loss:

$$\mathscr{L}_{Naive} = \frac{1}{|\{(u,i) : o_{u,i} = 1\}|} \sum_{(u,i):o_{u,i}=1} \delta_{u,i}(Y, \hat{Y})$$

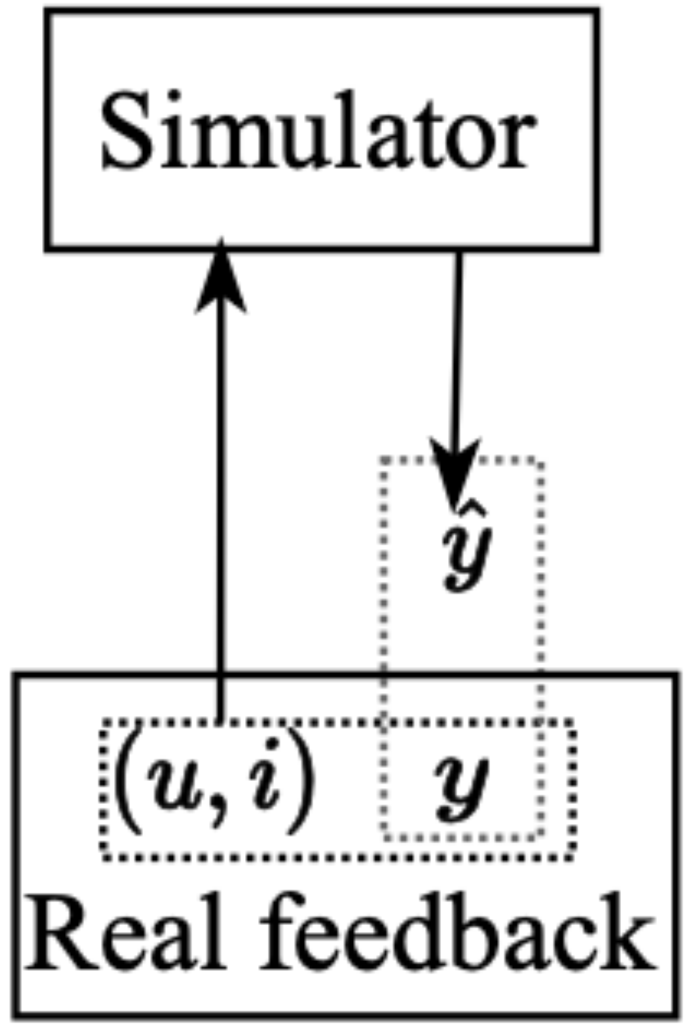
- Naive estimator

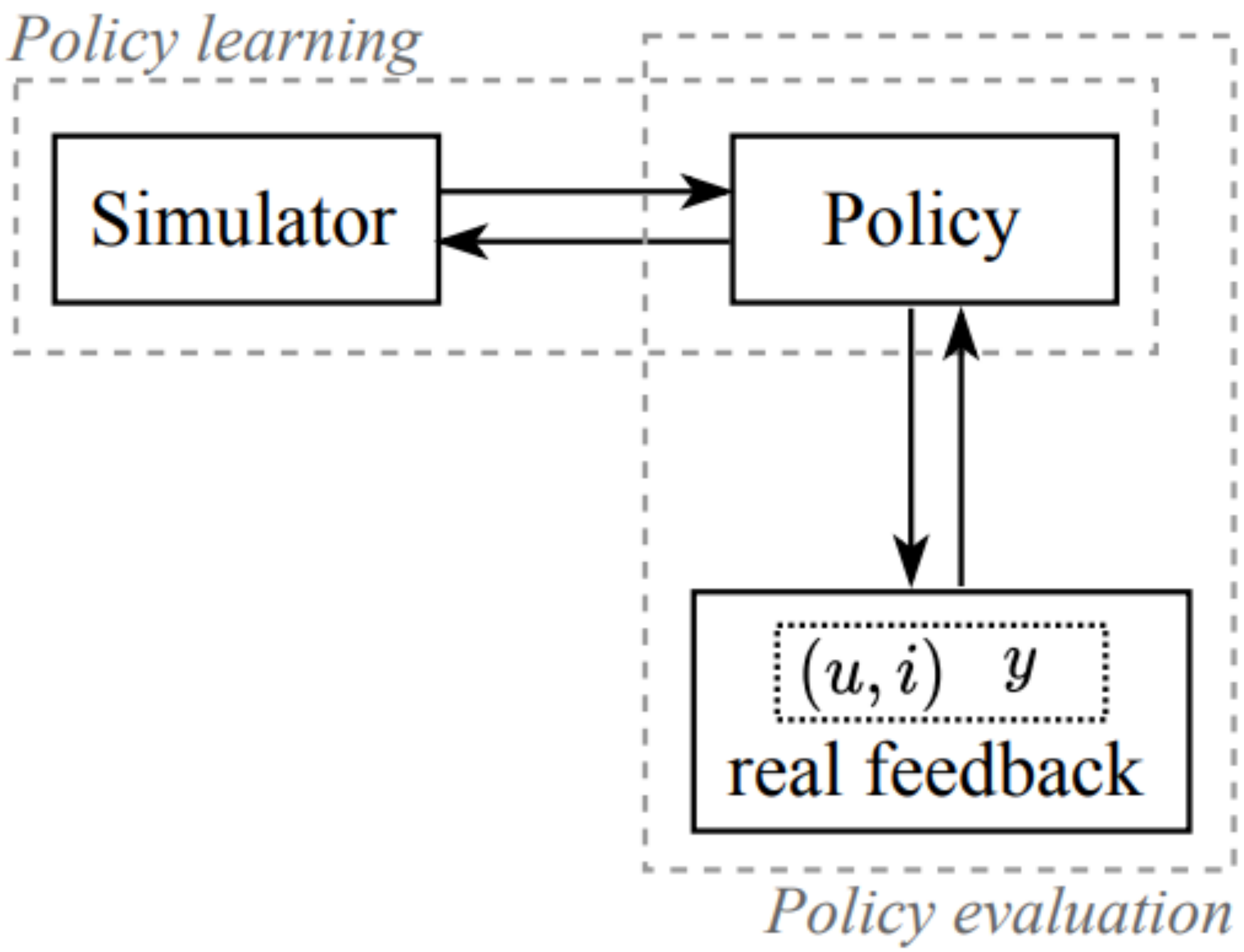
$$E[\mathscr{L}_{Naive}] = \frac{1}{\sum_{u=1}^{N} \sum_{i=1}^{M} P(o_{u,i} = 1)} \sum_{u=1}^{N} \sum_{i=1}^{M} P(o_{u,i} = 1)\delta_{u,i}(Y, \hat{Y}) \neq \frac{1}{N \cdot M} \sum_{u=1}^{N} \sum_{i=1}^{M} \delta_{u,i}(Y, \hat{Y})$$

- IPS-based estimator

$$E[\mathscr{L}_{IPS}] = \frac{1}{N \cdot M} \sum_{u=1}^{N} \sum_{i=1}^{M} \frac{P(o_{u,i} = 1)\delta_{u,i}(Y, \hat{Y})}{P(o_{u,i} = 1)} = \frac{1}{N \cdot M} \sum_{u=1}^{N} \sum_{i=1}^{M} \delta_{u,i}(Y, \hat{Y})$$

Evaluating the effect of bias in a simulation





- Evaluation based on observed user behavior
- Evaluation with considering the performance of policies

- Offline evaluation method

(Only requiring a sparse set of Missing-Completely-At-Random (MCAR) ratings)

(i) Train a policy using a simulator <u>with IBMS</u> on MNAR (***debiased policy***)

(ii) Train a policy using a simulator <u>without IBMS</u> on MNAR (***biased policy***)

(iii) Create a simulator based on MCAR ratings (***unbiased simulator***)

(iv) Deploy ***debiased*** *and* ***biased policies*** in the ***unbiased simulator***

- Solutions to the sparsity of MCAR ratings

Our simulator SOFA

- Two components:
  - **Debiased user-item rating matrix**

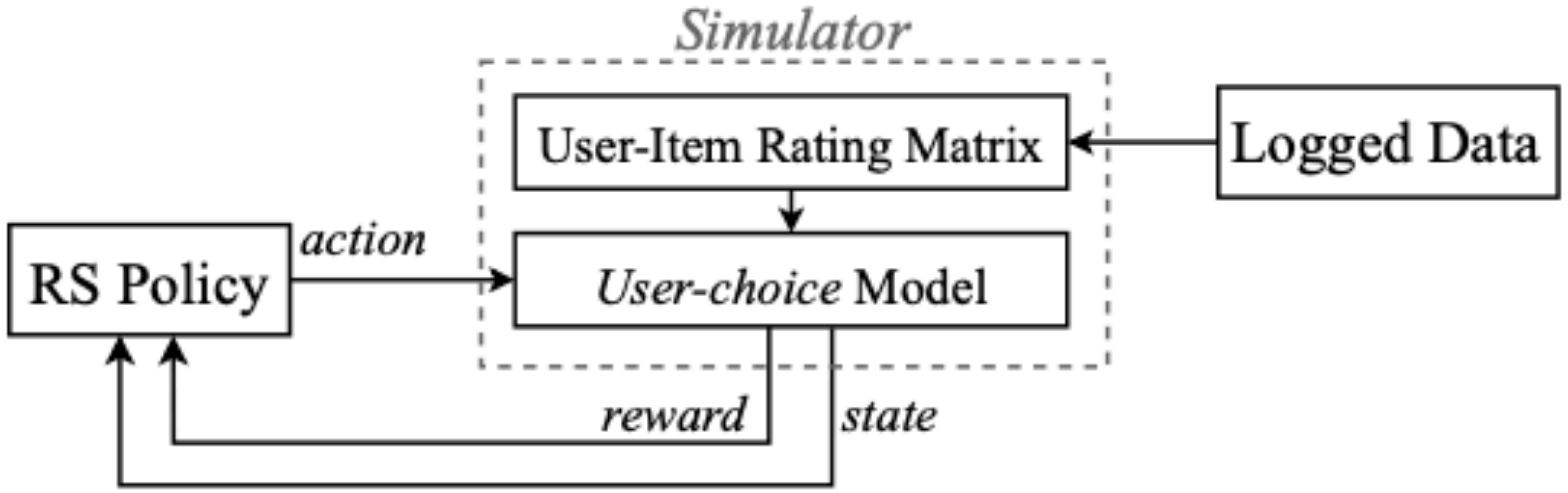
Produced using the IBMS with applying MF-IPS

  - **User-choice model**

(i) Feedback simulation: *click/skip*

(ii) State transition

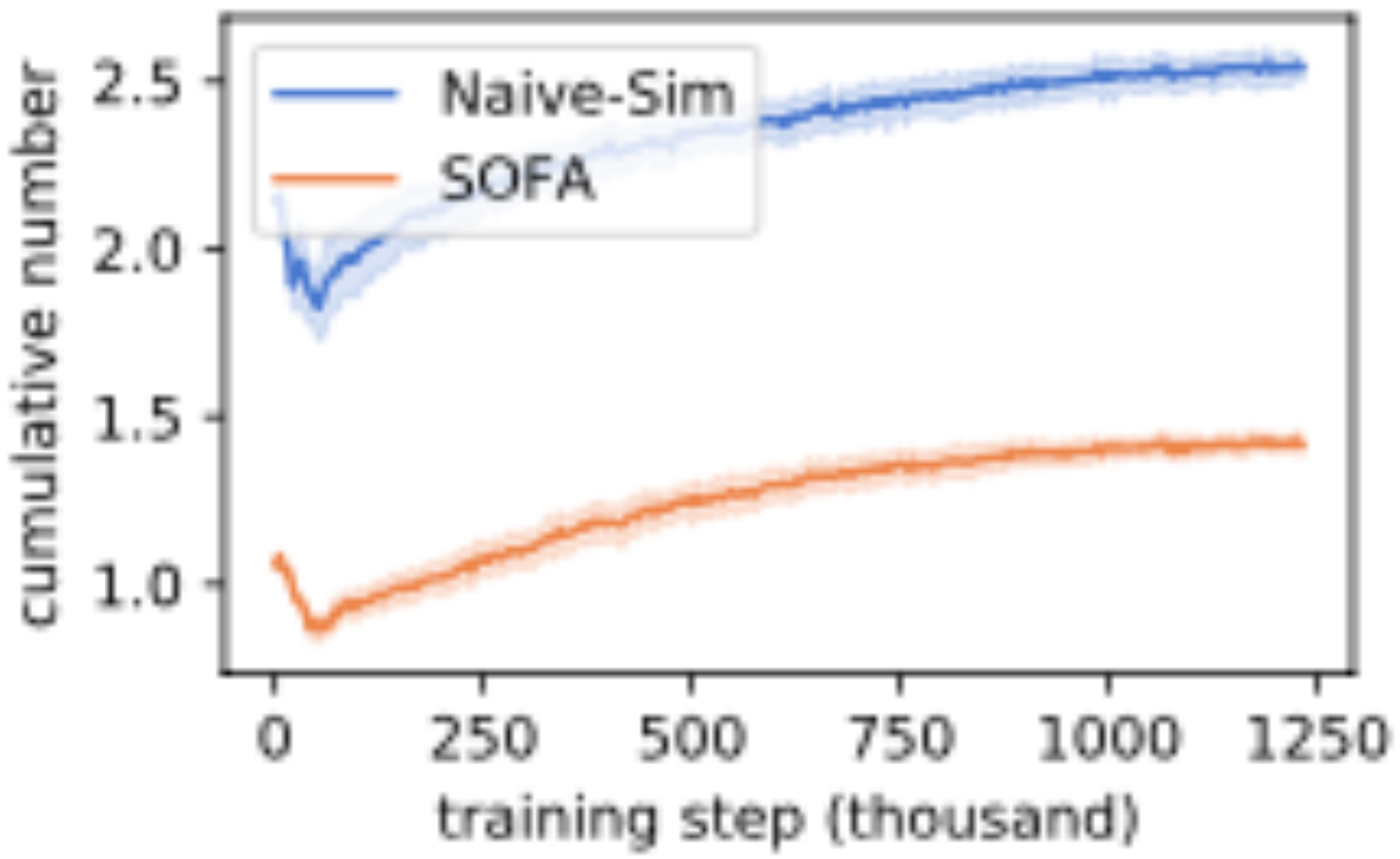
(iii) Reward generation



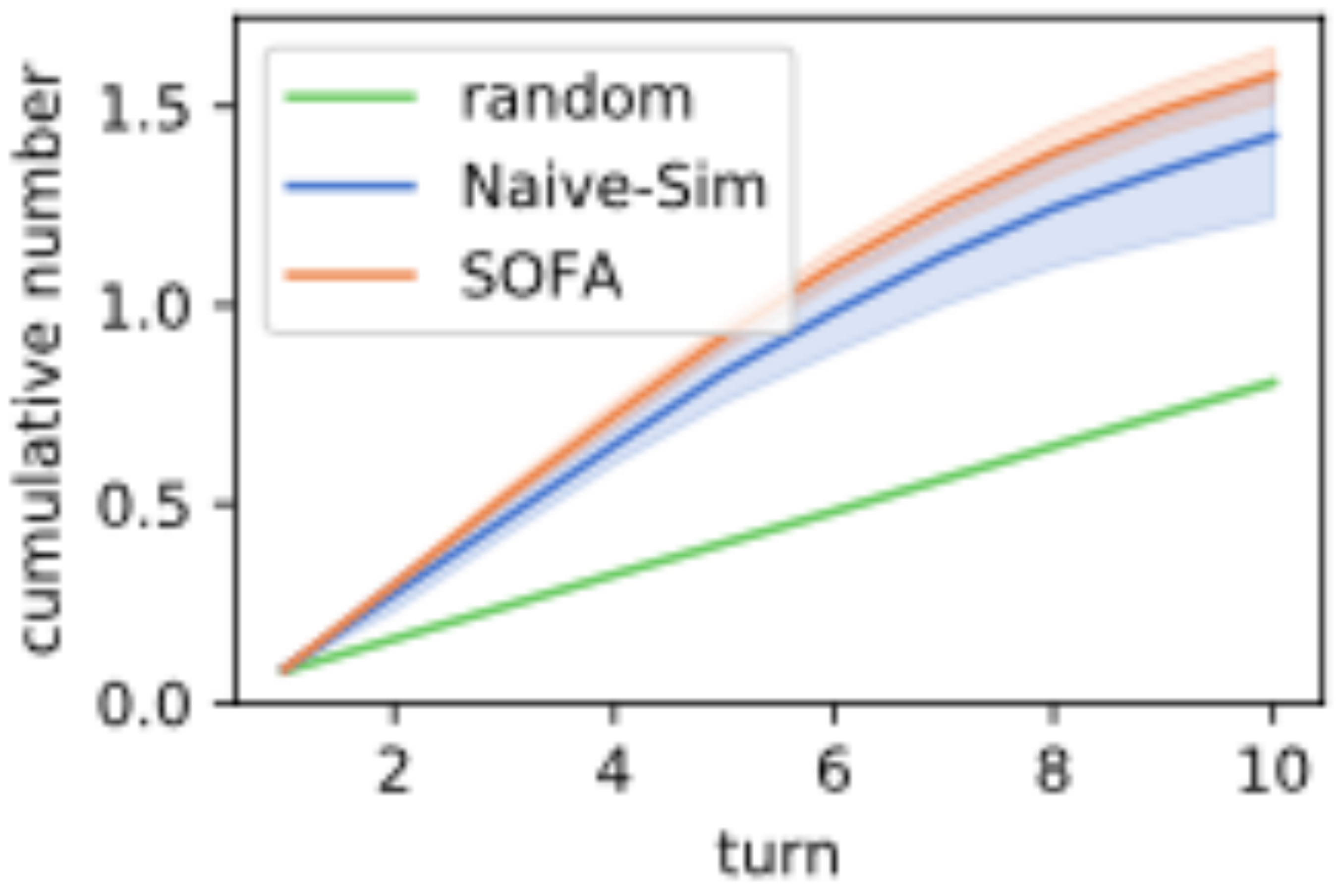
(b) RL4Rec with a simulator.

Experimental results

Training Curves



Evaluation results



(RQ1) Does interaction bias in logged data affect a simulator?

(RQ2) Can IBMS mitigate this bias effectively?

Conclusion

- Analysis on the effect of bias on RL4Rec simulators and the produced policies.
- *Intermediate Bias Mitigation Step* (IBMS) mitigates effect of bias.
- A novel way of evaluating the effect of bias on the final policy.
- SOFA, the first simulator for RL4Rec with correcting for bias.

Future work

- The effect of interaction bias on simulators for **multi-item recommendation** scenario.
- More widely relevant recommendation task such as **ranking with using implicit feedback**.

Thanks for listening