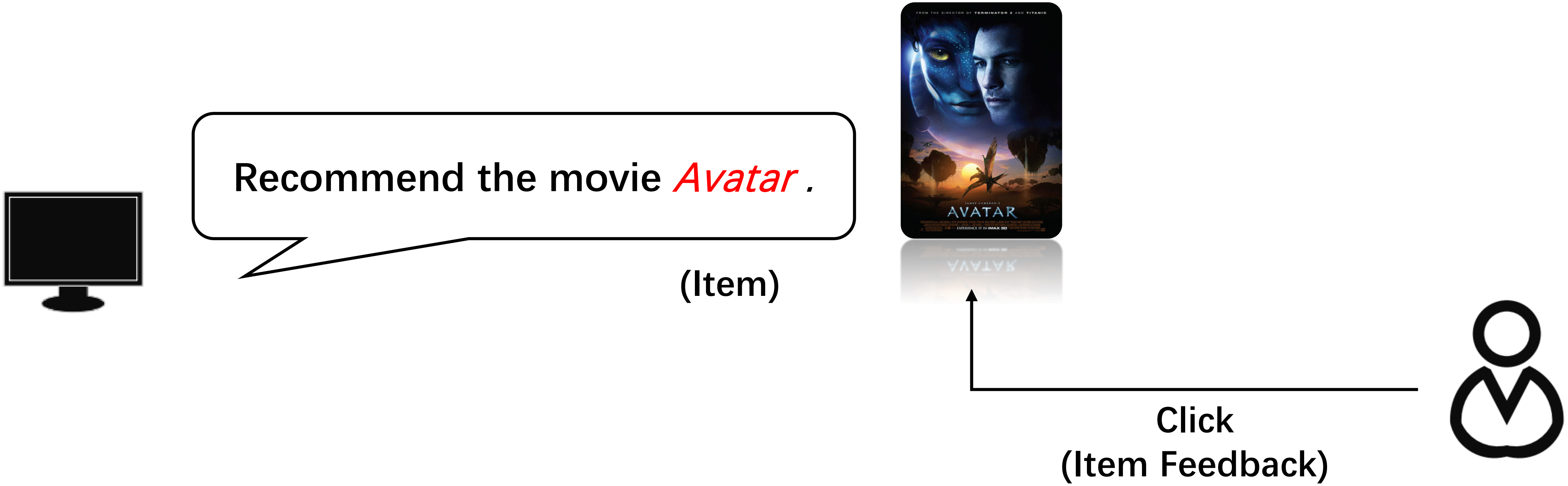
User-Regulation Deconfounded Conversational Recommender System with Bandit Feedback

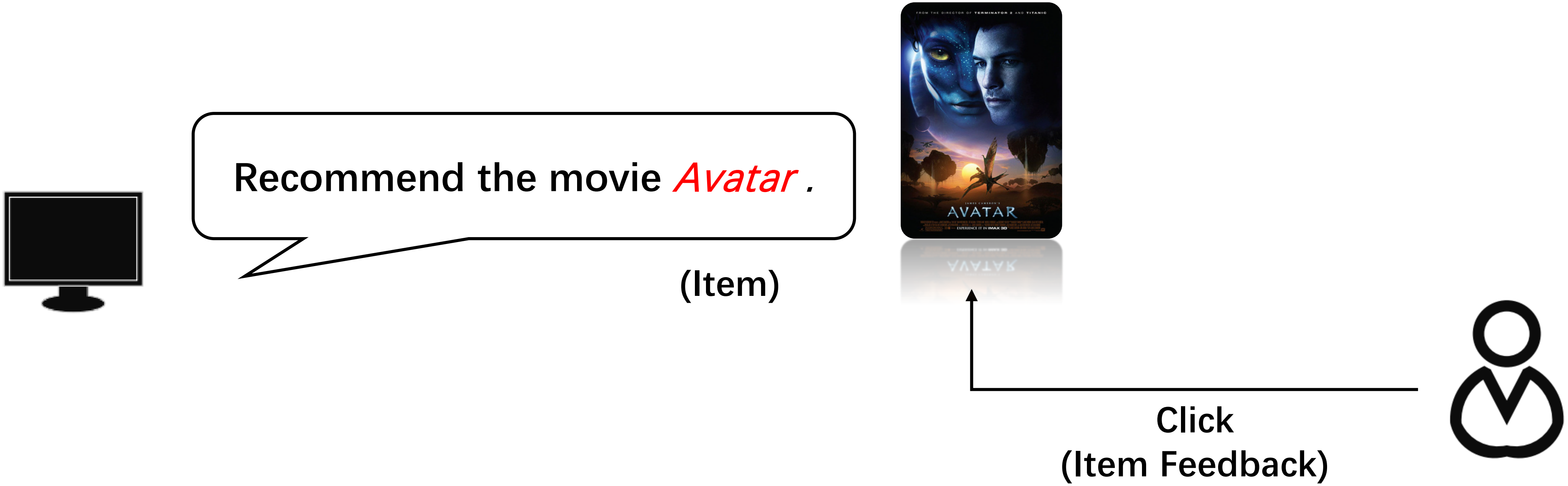
Yu Xia[1,4], Junda Wu[2], Tong Yu[3*], Sungchul Kim[3], Ryan A. Rossi[3], Shuai Li[1]
[1]Shanghai Jiao Tong University, [2]New York University, [3*]Adobe Research, [4]University of Michigan

Traditional Recommender System



Item Recommendation

Traditional Recommender System

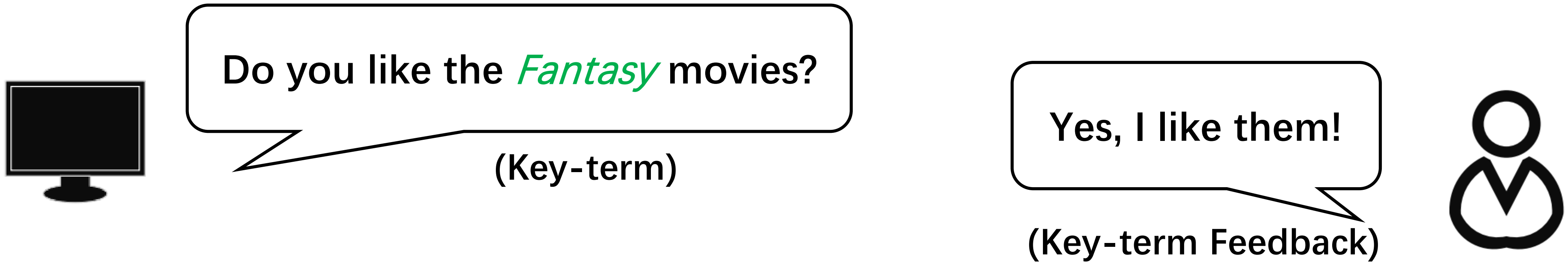


Item Recommendation

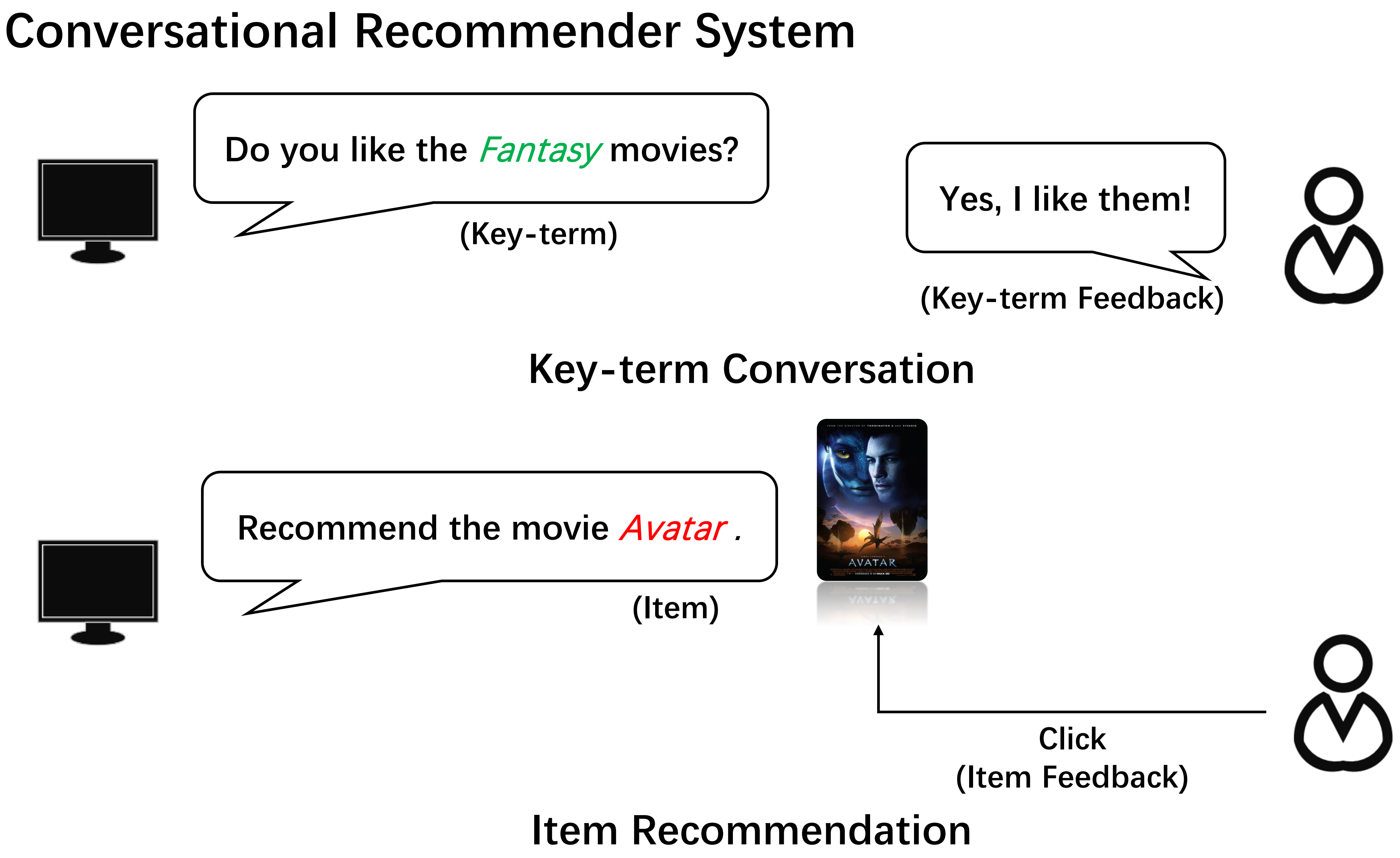
What if there is a **new** user?

How to deal with the **cold-start** scenario?

Conversational Recommender System



Key-term Conversation

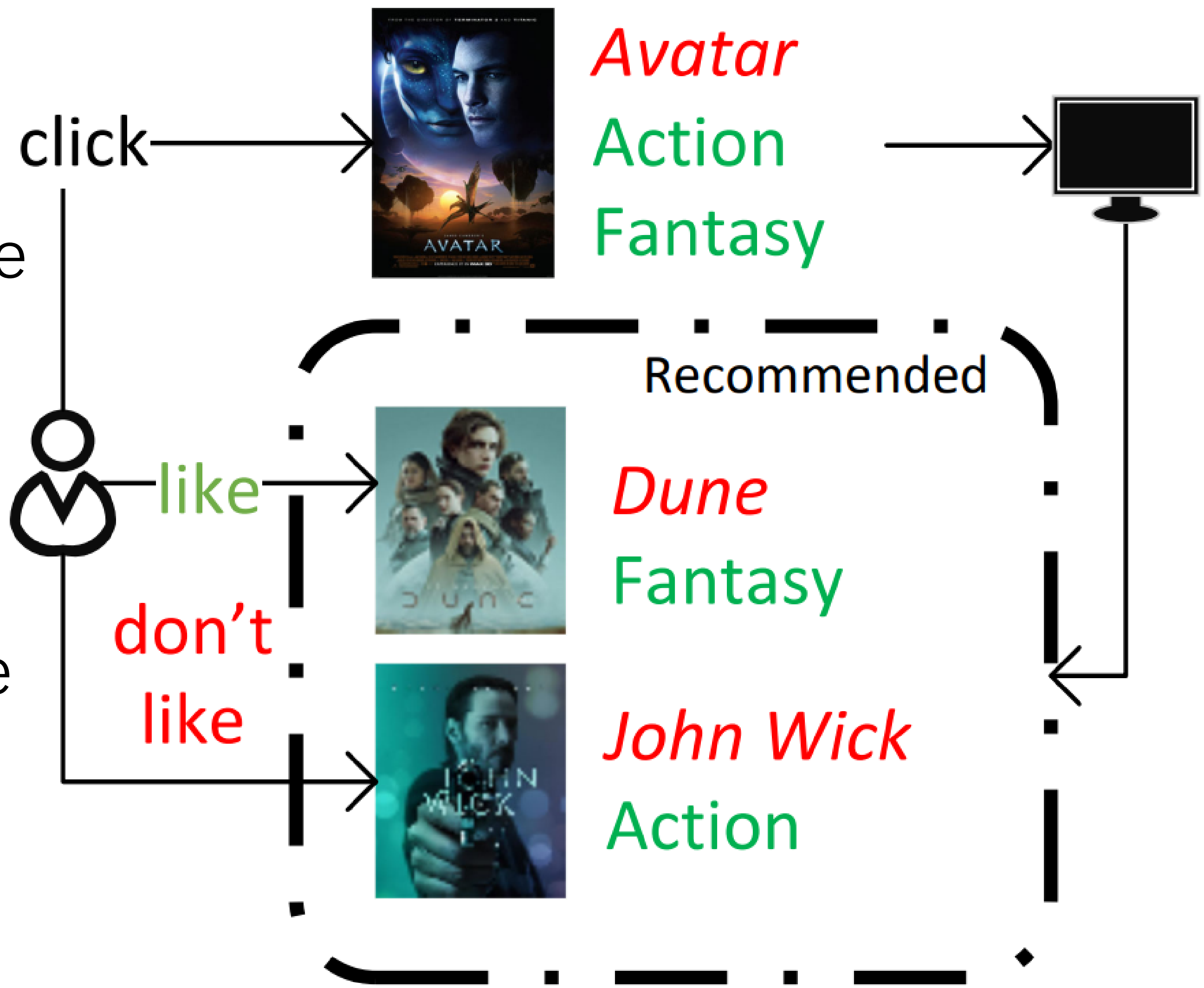


Conversational Recommender System
Do you like the Fantasy movies?
(Key-term)
Yes, I like them!
(Key-term Feedback)
Key-term Conversation
Recommend the movie Avatar .
(Item)
AVATAR
Click
(Item Feedback)
Item Recommendation

Conversational Recommender System

- The introduction of key-term conversation aims to accelerate the user preference learning.

- However, the explicit modelling of key-term preference can also introduce biases.



Causal View of CRS

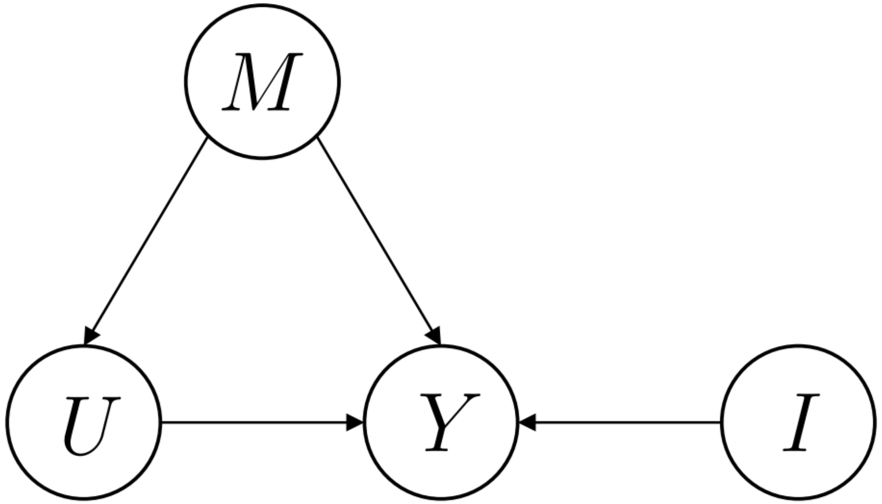
U: **item-level user preference** M: **key-term-level user preference**

I: **item representation** Y: **prediction score** K: **user regulation**

1. Conventional CRS

- Confounder: key-term-level user preference

$$
\begin{aligned}
&P(Y \mid U = u, I = i) \\
&\overset{(1)}{=} \sum_{m \in \mathcal{M}} P(Y, m \mid u, i) \\
&\overset{(2)}{=} \sum_{m \in \mathcal{M}} P(Y \mid u, i, m) P(m \mid u, i) \\
&\overset{(3)}{=} \sum_{m \in \mathcal{M}} P(Y \mid u, i, m) P(m \mid u)
\end{aligned}
$$



Causal View of CRS

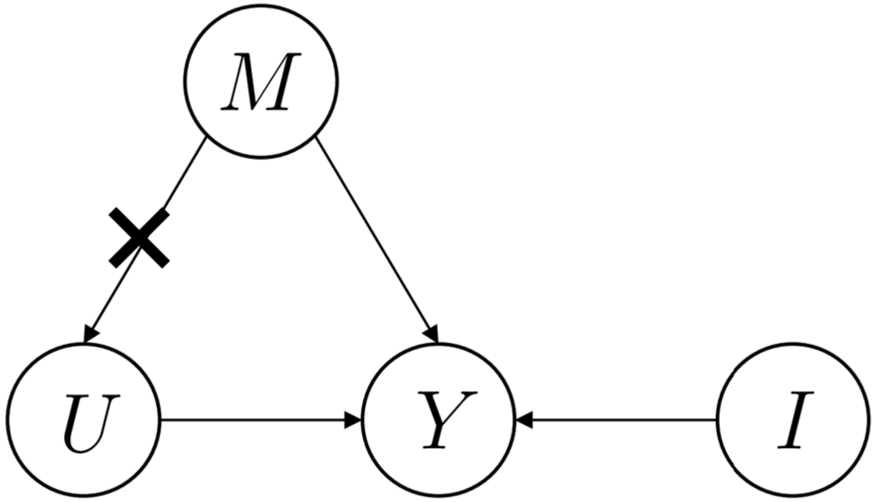
U: **item-level user preference** M: **key-term-level user preference**

I: **item representation** Y: **prediction score** K: **user regulation**

2. Deconfounded CRS w. Backdoor Adjustment

- *do*-calculus eliminating confounding effect

$$
\begin{aligned}
&P(Y \mid do(U=u), I=i) \\
&\overset{(1)}{=} \sum_{m \in \mathcal{M}} P(Y \mid do(U=u), i, m) P(m \mid do(U=u), i) \\
&\overset{(2)}{=} \sum_{m \in \mathcal{M}} P(Y \mid do(U=u), i, m) P(m \mid do(U=u)) \\
&\overset{(3)}{=} \sum_{m \in \mathcal{M}} P(Y \mid u, i, m) P(m).
\end{aligned}
$$



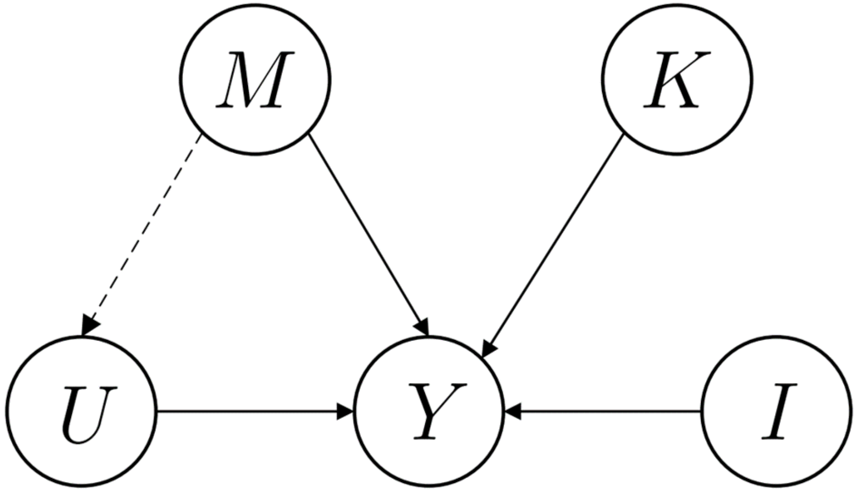
Causal View of CRS

U: item-level user preference **M: key-term-level user preference**

I: item representation **Y: prediction score** **K: user regulation**

3. Deconfounded CRS w. User-Regulated Backdoor Adjustment

- Not all biases are harmful.
- Utilize user feedback to identify useful biases
- Adaptively regulate the effect of backdoor adjustment.



User-Regulation Coefficient

$$\hat{P}(Y \mid U, I) = \eta \underbrace{P(Y \mid U, I)}_{\text{Before Backdoor Adjustment}} + (1-\eta) \underbrace{P(Y \mid do(U), I)}_{\text{After Backdoor Adjustment}}$$

User-Regulation Deconfounded CRS

I recommend the movie *Avatar* because you might like *Action* movies.

Well, I love *Avatar*, but I am not a big fan of *Action* movies.

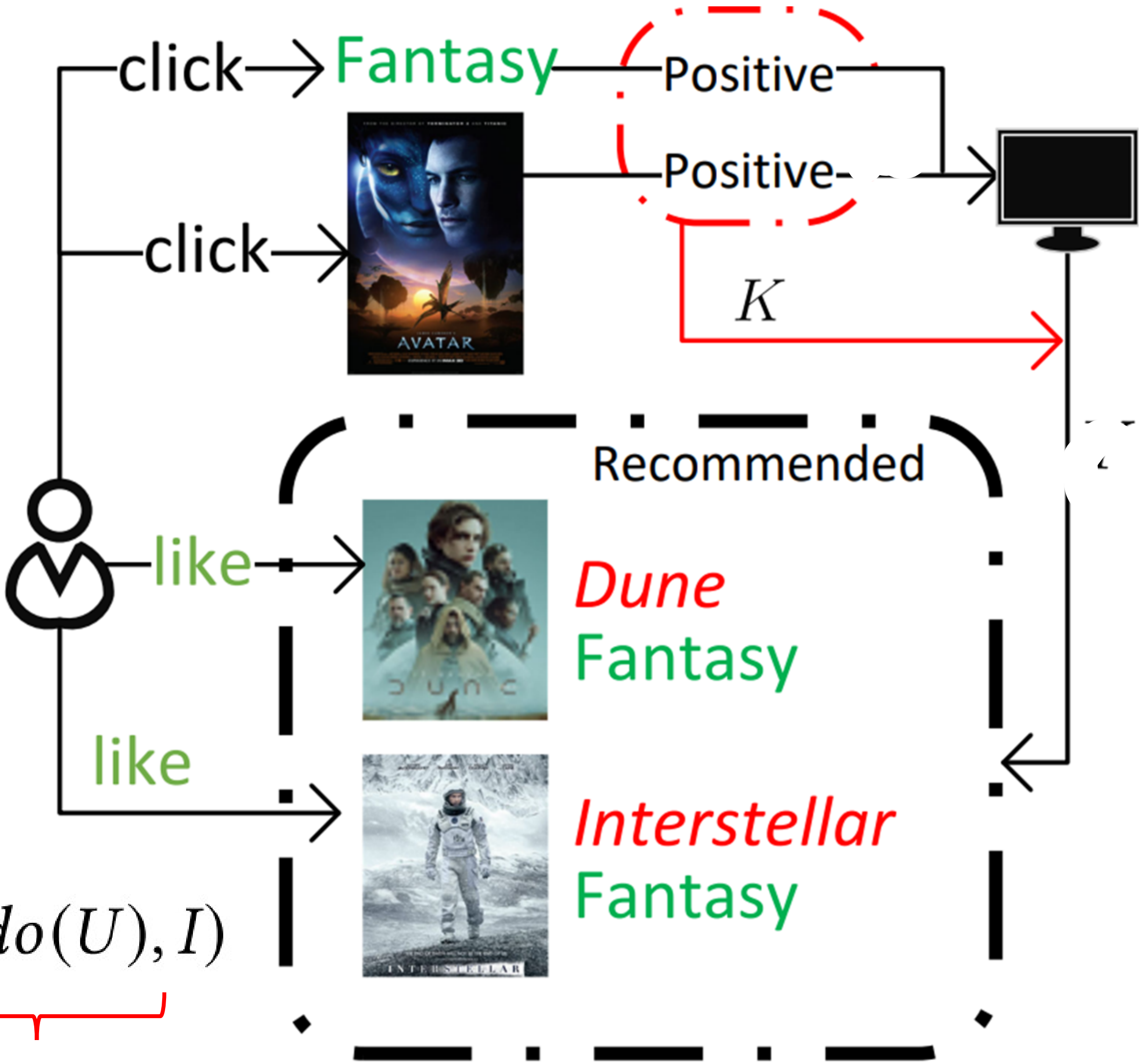
From the item feedback together with the key-term feedback, we can potentially capture more accurate preference relations

User-Regulation Deconfounded CRS

- User-Regulation Coefficient

$$\eta = \frac{\sum_{\mathcal{D}} \mathbf{1}[(Positive, Positive)]}{\sum_{\mathcal{D}} \mathbf{1}}$$

- Adaptively control the effect of backdoor adjustment in a personalized manner.



$$\hat{P}(Y \mid U, I) = \eta \underbrace{P(Y \mid U, I)}_{\text{Before Backdoor Adjustment}} + (1-\eta) \underbrace{P(Y \mid do(U), I)}_{\text{After Backdoor Adjustment}}$$

User-Regulation Deconfounded CRS (Implementation)

- **DecUCB:** A new upper-confidence-based contextual bandit algorithm for CRS.
- **EE Trade-off:** Balance **E**xploration and **E**xploitation in the cold-start setting.
- **Online Debiasing:** User-regulated backdoor adjustment in an online algorithm.

For more technical details, please refer to our paper.

Algorithm 1: DecUCB

Input: 1) user $u \in \mathcal{U}$ with $t = t_u$ rounds of historical interaction, parameters $\mathbf{M}_u$, $\boldsymbol{b}_u$, $\tilde{\mathbf{M}}_u$ and $\tilde{\boldsymbol{b}}_u$;
2) candidate arms $\mathcal{A}_t$, related key-terms $\mathcal{K}_t$, weight $\mathbf{W}$;
3) regularization parameters λ, $\tilde{\lambda}$, and exploration parameters α, $\tilde{\alpha}$, user regulation scale ϵ.

```
if t = 0 then
    Initialize M_u = (1 − λ)I, b_u = 0, M̃_u = λ̃I, b̃_u = 0;
Compute θ̃_u = M̃_u^{-1} b̃_u;
Compute θ according to Section 5.2;
Select an item a_t ∈ A_t according to Equation 14;
Collect the set of key-terms K_{a_t} related to a_t;
Select an key-term k_t ∈ K_{a_t} according to Section 5.1 ;
Receive the item feedback r_{a_t,t} and the key-term feedback
  r̃_{k_t,t} according to Equation 1 and Equation 3, respectively;
Update M_u = M_u + λ x_{a_t} x_{a_t}^⊤, b_u = b_u + λ r_{a_t} x_{a_t};
Update M̃_u and b̃_u according to Algorithm 2;
```

User-Regulation Deconfounded CRS (Insights)

- **Item Selection:** Based on the deconfounded reward estimation plus the corresponding upper-confidence bound.

$$a_t = \arg\max_{a \in \mathcal{A}_t} Y_{a,u} + \lambda\alpha \left\|\boldsymbol{x}_a\right\|_{\mathbf{M}_u^{-1}} + (1-\lambda)\tilde{\alpha} \left\|\mathbf{M}_u^{-1}\boldsymbol{x}_a\right\|_{\tilde{\mathbf{M}}_u^{-1}}$$

- **Key-term Selection:** Chosen from key-terms related to the recommended item, which reduces the most uncertainty in the next round of recommendation.

$$k = \arg\max_{k' \in \mathcal{K}_{a_t}} \frac{\left\|\mathbf{X}_t \mathbf{M}_u^{-1} \tilde{\mathbf{M}}_u^{-1} \tilde{\boldsymbol{x}}_{k'}\right\|_2^2}{1 + \tilde{\boldsymbol{x}}_{k'}^{\top} \tilde{\mathbf{M}}_u^{-1} \tilde{\boldsymbol{x}}_{k'}},$$

For more technical details, please refer to our paper.

Experiments

- **Datasets**

Table 1: Statistics of the datasets used in our experiments.

Dataset	Last.FM	MovieLens	BibSonomy
# of users	400	400	400
# of items	2,000	2,000	2,000
# of key-terms	2,726	5,585	5,472
avg. # of related key-terms	16.55	11.37	16.75
avg. # of related items	12.14	4.071	6.12

- **Baselines**

  - **LinUCB** [15]: A state-of-the-art contextual bandit approach without conversational feedback.
  - **ConUCB** [40]: A recently proposed conversational contextual bandit algorithm that conducts key-term conversations with users before recommending an item. For fair comparison, we enable ConUCB to conduct key-term conversation in each round.
  - **$\mathbf{DecUCB}_{\mathbf{w/o\ UR}}$**: A variant of our proposed DecUCB that performs backdoor adjustment without any user regulation.

Experiments

- Recommendation Accuracy (Cumulative Regret)

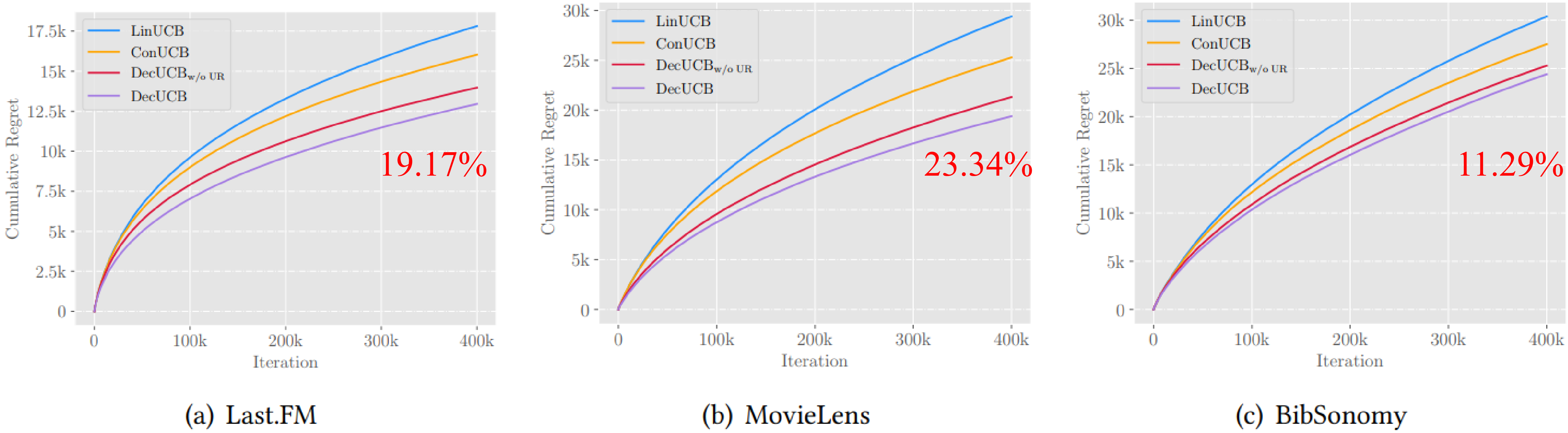


Figure 3: Cumulative regret on Last.FM, MovieLens and BibSonomy datasets with equal number of user feedback.

Experiments

- Bias Mitigation (KL-Divergence)

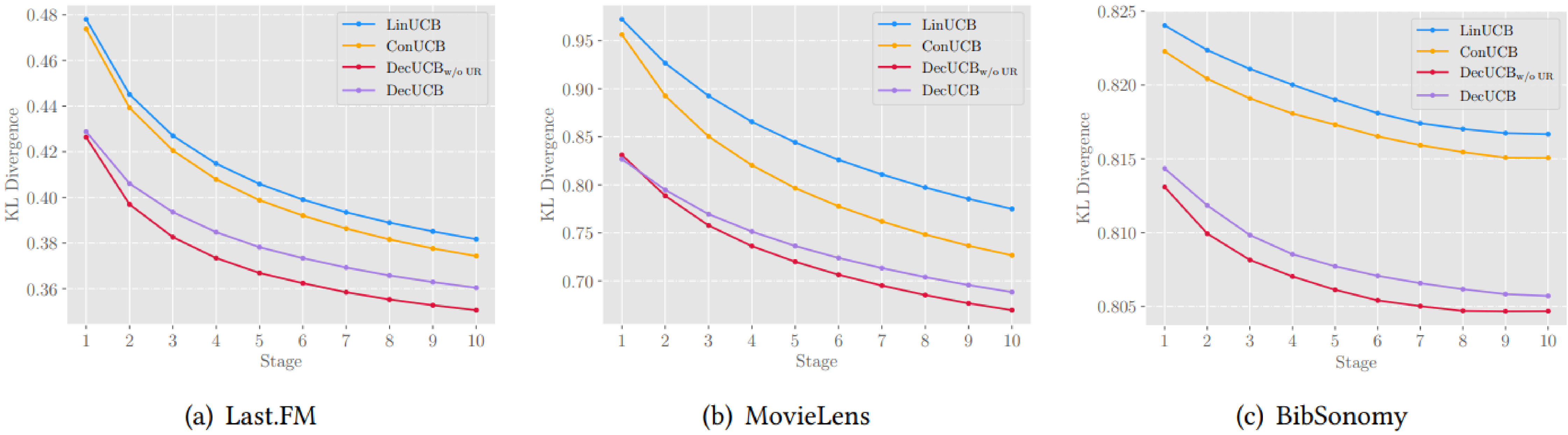
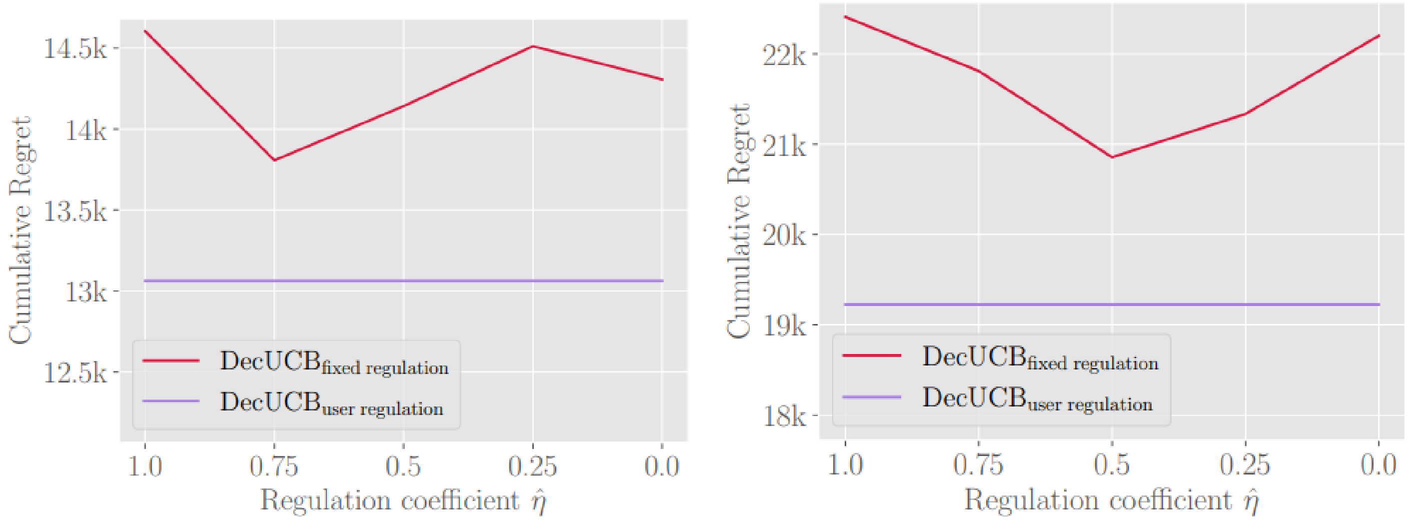


Figure 4: KL divergence across difference stages of recommendation with equal number of user feedback.

Experiments

- User-Regulation Coefficient Analysis



(a) Last.FM (b) Movielens

Figure 6: Effect of different regulation coefficient $\hat{\eta}$.

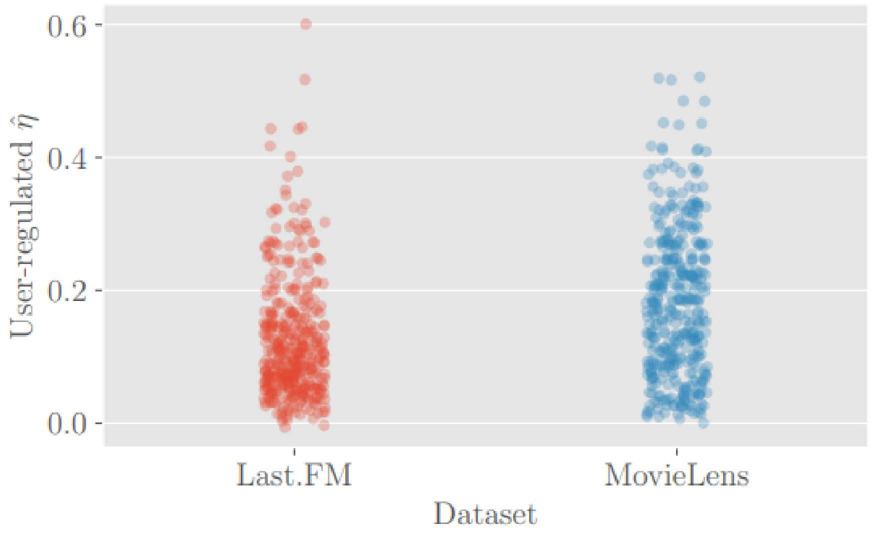
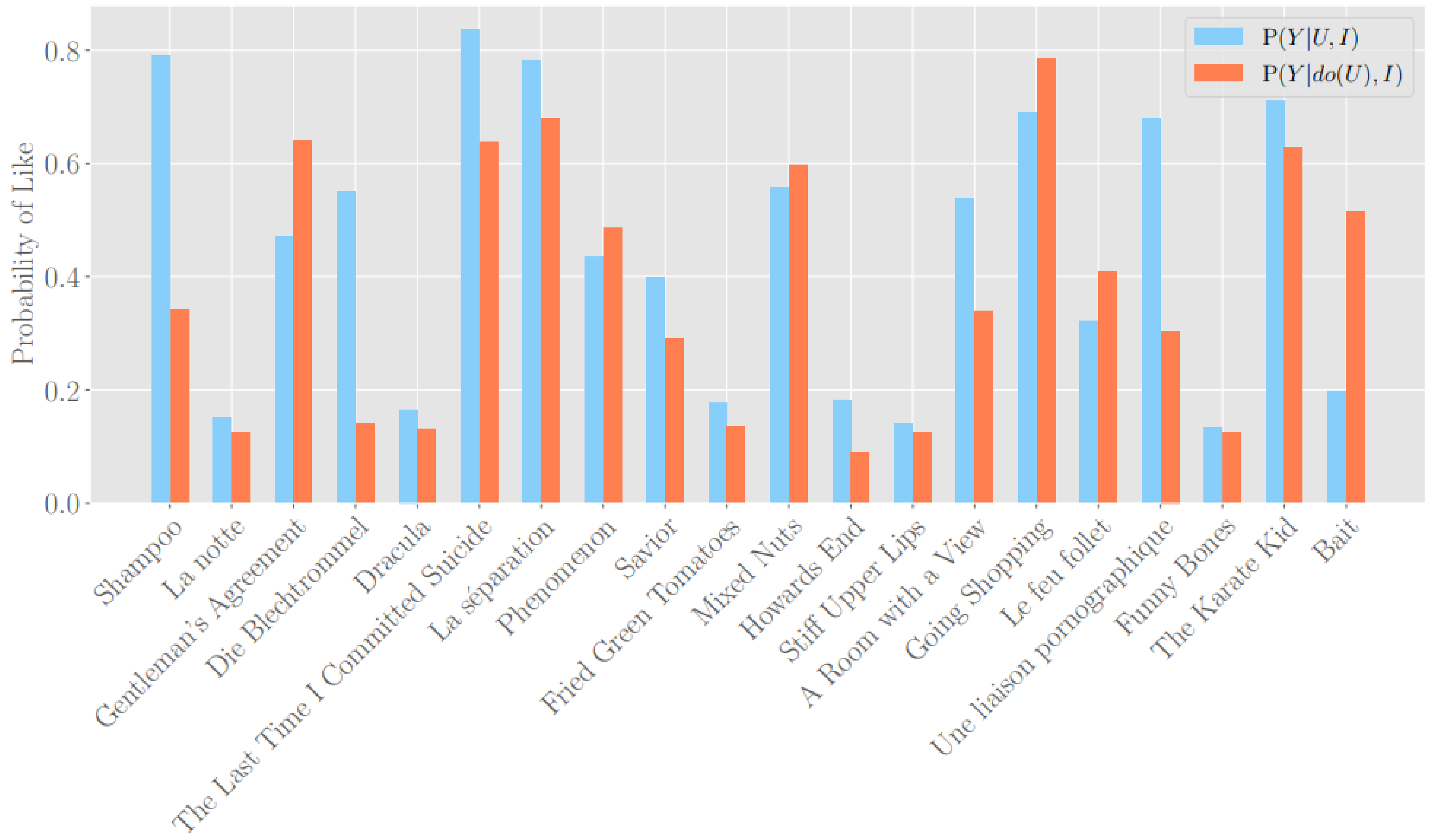


Figure 7: Visualization of user regulation coefficient $\hat{\eta}$ on Last.FM and MovieLens dataset.

Experiments

Figure 5: Prediction scores of movies in MovieLens dataset before and after *do*-calculus of backdoor adjustment.



Conclusion

- We studied the biases in CRS from a causal perspective.
- We proposed a Deconfounded CRS that effectively alleviate biases.
- We enable users to provide feedback to adaptively regulate debiasing.

Future Work

- More fine-grained user feedback for more accurate bias identification.
- Explainable CRS modelling more accurate causal relations.

Contact: xiayu9@sjtu.edu.cn