Causal Recommendation: Progresses and Future Directions

Lecture Tutorial for SIGIR 2023

Organizers: Wenjie Wang, Yang Zhang, Haoxuan Li, Peng Wu, Fuli Feng, Xiangnan He

23 July 2023

Webpage: https://causalrec.github.io/

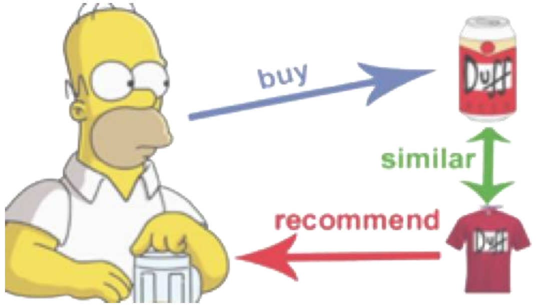
Outline

- Part 1 (90 min, 9:00—10:30)
  - Introduction (Wenjie Wang, 15 min)
  - Structural causal models for recommendation (Yang Zhang and Wenjie Wang, 60~70 min)
  - Q&A (5 min)
  - Coffee break (30 min)
- Part 2 (90 min, 11:00-12:30)
  - Potential outcome framework for recommendation (Haoxuan Li and Peng Wu, 60~70 min)
  - Comparison (Fuli Feng, 2 min)
  - Conclusion, open problems, and future directions (Fuli Feng, 20 min)
  - Q&A (5 min)

Information Seeking

- **Information explosion era**
  - E-commerce: 12 million items in Amazon.
  - Social networks: 2.8 billion users in Facebook.
  - Content sharing platforms: 720,000 hours videos uploaded to Youtube per day.

- **Recommender system**



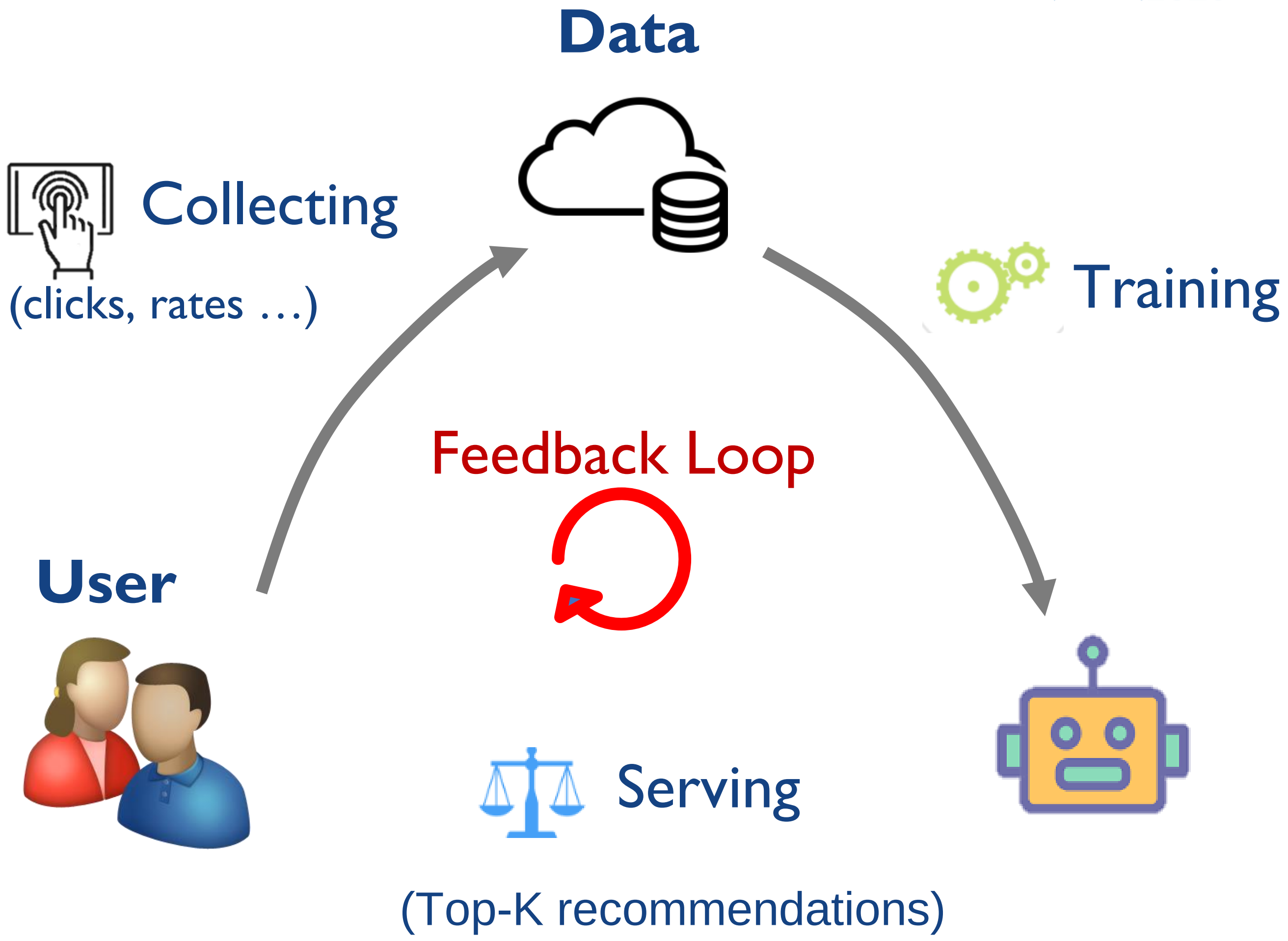
Recommendation

Information seeking
via **implicit feedback**

Recommender system is a powerful tool to address information overload.

Ecosystem of RecSys

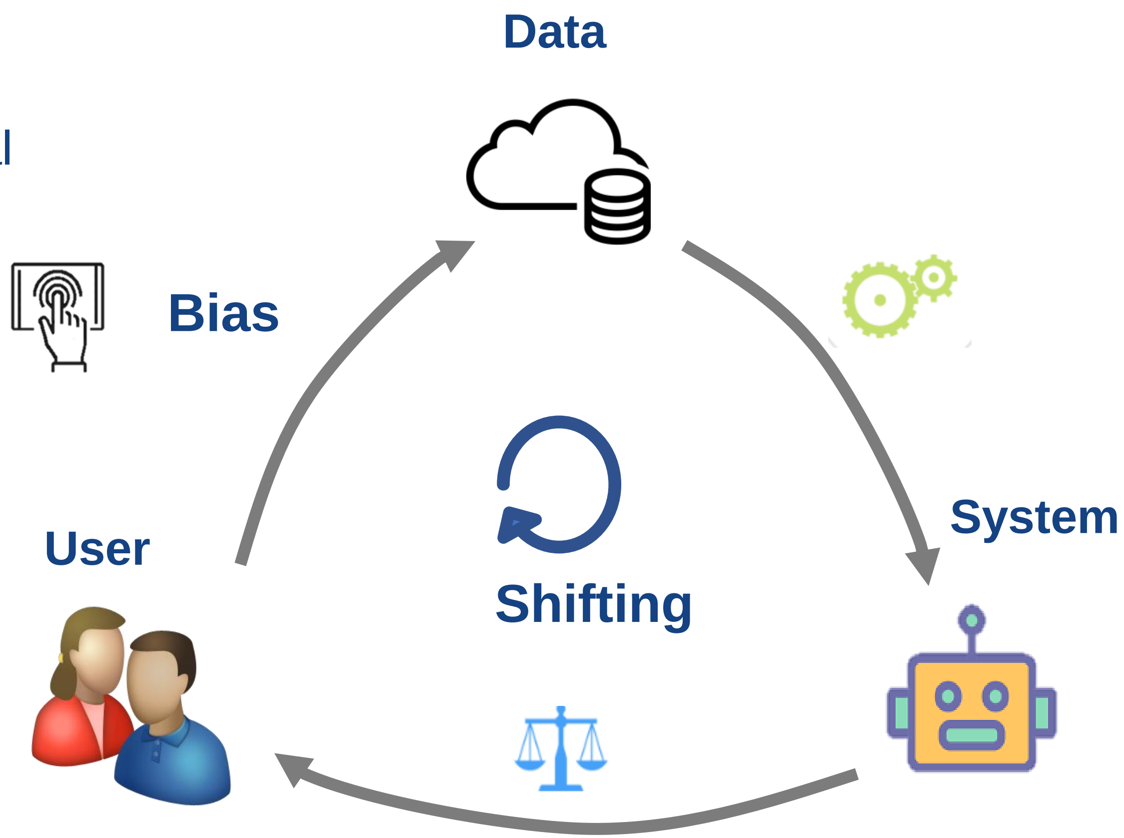
- Workflow of RecSys
  - **Training**: RecSys is trained on observed user-item interactions.
  - **Serving**: RecSys infers user preference over items and recommend Top-K items.
  - **Collecting**: collect user interactions on the recommended items for further training.
- Forming a feedback loop



Chen et al. arxiv 2021. Bias and Debias in Recommender System: A Survey and Future Directions

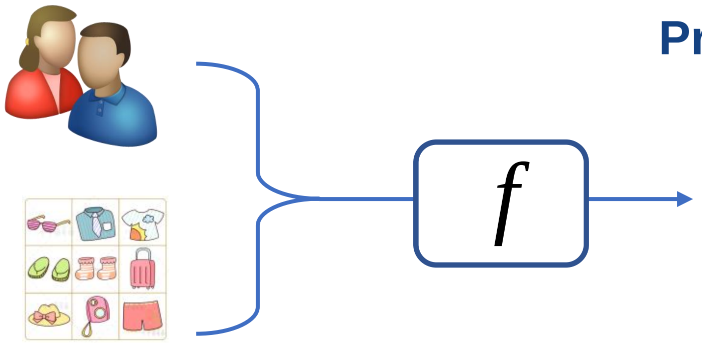
Shortcomings of Data-driven RecSys

- **Bias in data (collecting):**
  - Data is observational rather than experimental (missing-not-at-random)
  - Affected by many hidden factors:
    - Public opinions, *etc.*
- **Bias shifting along time:**
  - User/item feature changes
    - Income, marriage status
  - Preference shifting



Fitting Historical Data

- Minimizing the difference between historical feedback and model prediction



Predicted Score

3	2	1	4
2	2	3	5
2	4	5	5

$\approx$

History feedback

3		1	
2	3		5
2		5	

- **Collaborative filtering:** Similar users perform similarly in future

Shallow representation learning

- Matrix factorization & factorization machines

	User				Movie					Other Movies rated					Time	Last Movie rated					Target y	
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...	Time	TI	NH	SW	ST	...		
$x^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y^{(1)}$
$x^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y^{(2)}$
$x^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y^{(2)}$
$x^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y^{(3)}$
$x^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y^{(4)}$
$x^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y^{(5)}$
$x^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y^{(6)}$

Feature vector **x**

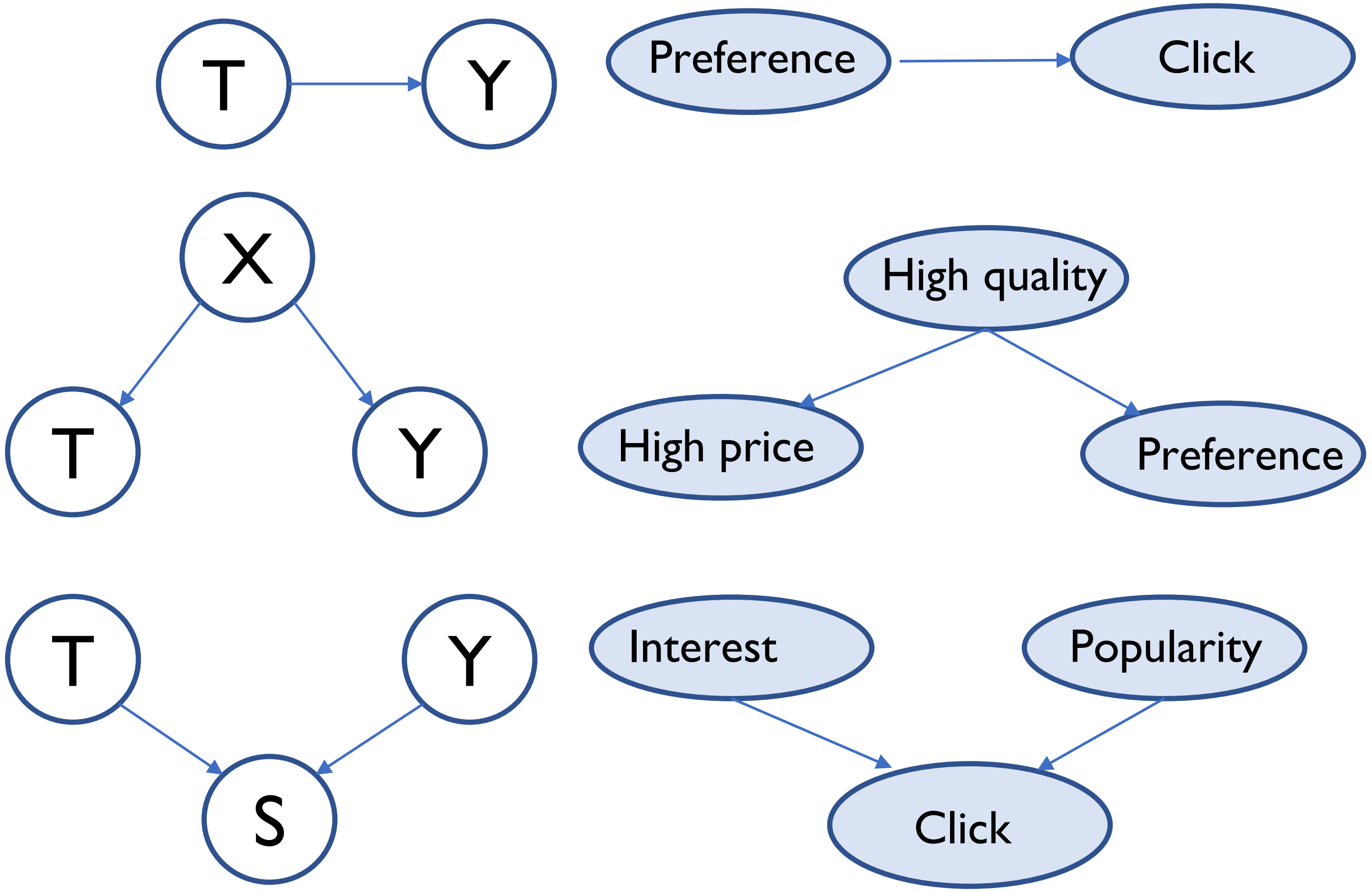
Neural representation learning

- Neural collaborative filtering
- Graph neural networks
- Sequential model
- Textual & Visual encoders

Learning correlations between input features and interactions.

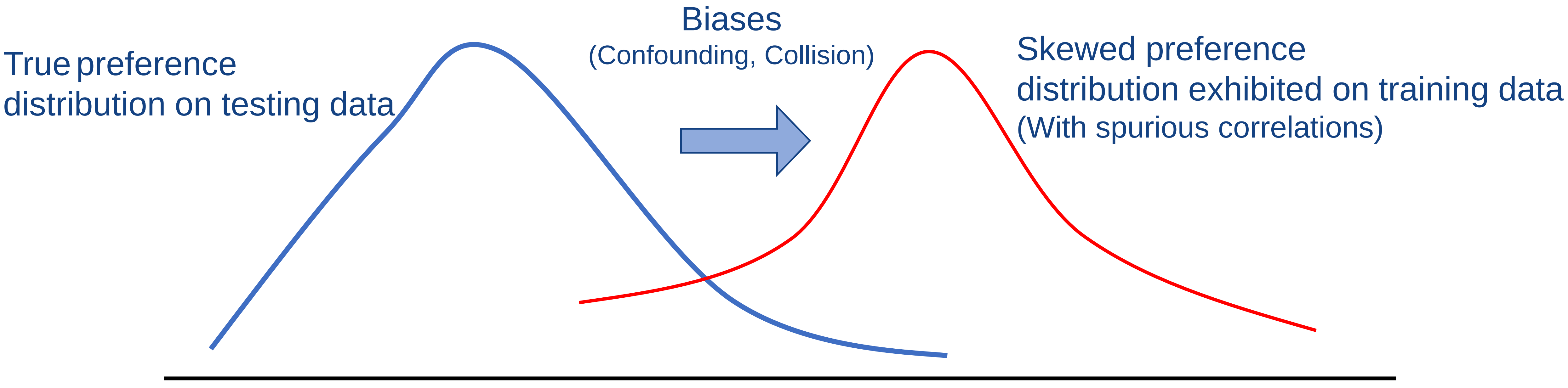
Shortcomings of Data-driven RecSys

- Correlation != preference: Correlations may not reflect the true causes of interactions.
- Three basic types of correlations:
  - **Causation**
    - Stable and explainable
  - **Confounding**
    - Ignoring X
    - Spurious correlation
  - **Collision**
    - Condition on S
    - Spurious correlation



Shortcomings of Data-driven RecSys

- Data-driven methods will learn skewed user preference:



- Data-driven methods may infer spurious correlations, which deviates from users' true preference.

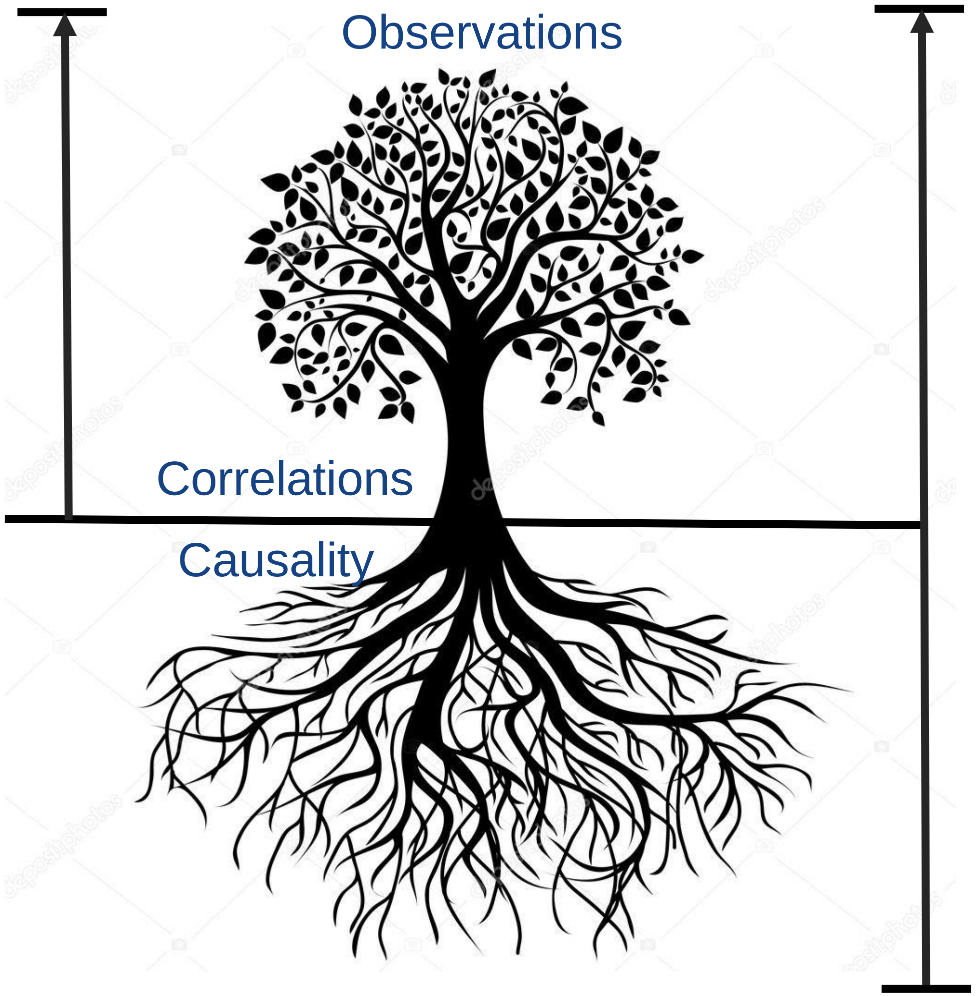
Correlation != preference

Chen et al. TOIS 2023. Bias and Debias in Recommender System: A Survey and Future Directions

Why Causal Inference?

- Aim: Understanding the inherent **causal mechanism** behind user behaviors
  - Capturing user true preference

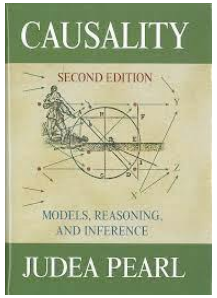
- Making reliable & explainable recommendations
  - Correlation + Causality > Correlation



Classification of Causal Recommendation

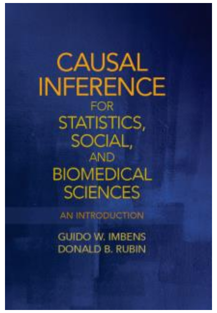
- Structural Causal Model (SCM)

(Judea Pearl)



- Potential Outcome Framework

(Donald B. Rubin)



Evaluation

Debiasing

Explanation

Recommendation

Fairness

Robustness & OOD generalization

Outline

- Part 1 (90 min, 9:00—10:30)
  - Introduction (Wenjie Wang, 15 min)
  - Structural causal models for recommendation (Yang Zhang and Wenjie Wang, 60~70 min)
  - Q&A (5 min)
  - Coffee break (30 min)
- Part 2 (90 min, 11:00-12:30)
  - Potential outcome framework for recommendation (Haoxuan Li and Peng Wu, 60~70 min)
  - Comparison (Fuli Feng, 2 min)
  - Conclusion, open problems, and future directions (Fuli Feng, 20 min)
  - Q&A (5 min)

Structural Causal Model

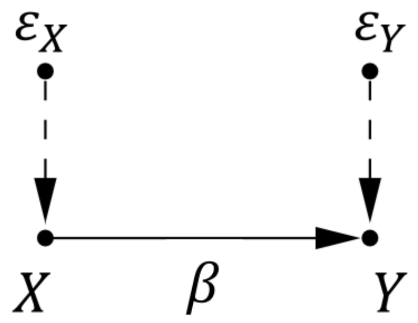
- How can common understandings, such as the fact that symptoms do not cause diseases, be expressed mathematically?

X: disease Y: *symptom*

$$X = U_X$$
$$Y = \beta X + U_Y$$

U_X *and* U_Y: exogenous

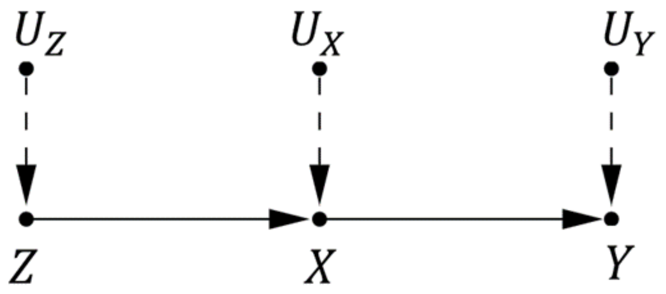
To express the inherent directionality



Causal Graph / Causal Diagram

Causal diagrams encodes causal assumption via missing arrows, representing claims of zero influences

- General form:



Non-parametric interpretation

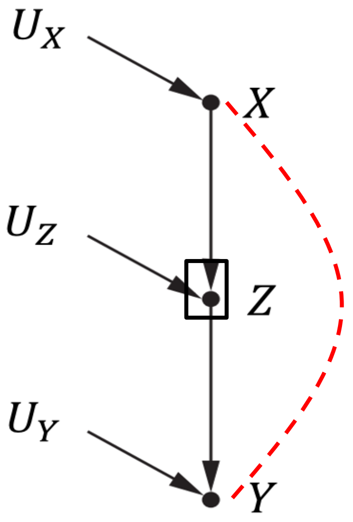
$$Z = f_Z(U_Z)$$
$$X = f_X(Z, U_X)$$
$$Y = f_Y(X, U_Y)$$

Pearl, Judea. "Causal inference." Causality: objectives and assessment (2010): 39-58.

Structural Causal Model

- Basic causal structures in causal graph

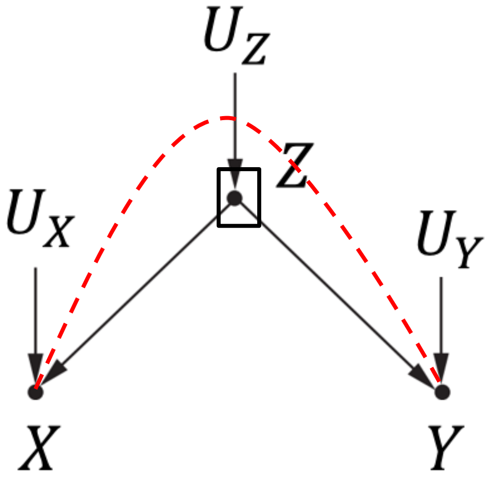
Chain



Z: **mediator**

- X and Y are associated.
- condition on Z, X and Y are independent.

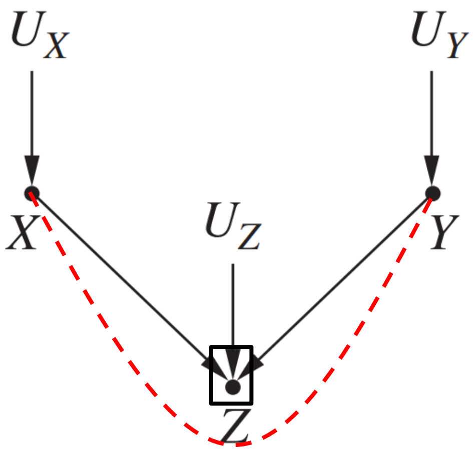
Confounding



Z: **confounder**

- X does not affect Y, but X and Y are correlated. (Spurious correlations).
- condition on Z, X and Y are independent, blocking the spurious correlations.

Colliding



Z: **collider**

- X and Y are independent.
- Condition on Z, X and Y are correlated, bringing spurious correlations.

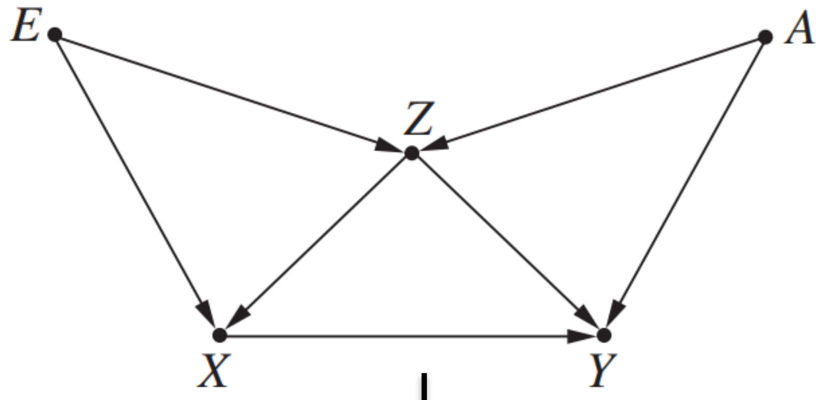
Structural Causal Model

- Correlation is not causation

  Confounders and controlling colliders would bring spurious correlations between treatment and outcome.

  It is impossible to answer causal question with correlation-level tools

Confounder E,Z,A will bring spurious correlations



$$P(Y|do(X)) = \sum_{z,a} P(Y|X,z,a)P(z,a)$$

- *do*-calculus

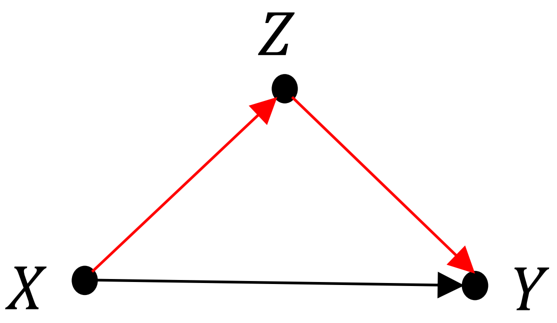
  It provides various principles to identify target causal effect.
  For example, utilize *the backdoor adjustment when confounders exist*

> *If any node in Z isn't a descendant of X, and Z blocks every path between X and Y that contains an arrow into X (backdoor path), then the average causal effect of X on Y is:*
> $$P(Y|do(X)) = \sum_Z P(Y|X,Z)P(Z)$$

Structural Causal Model

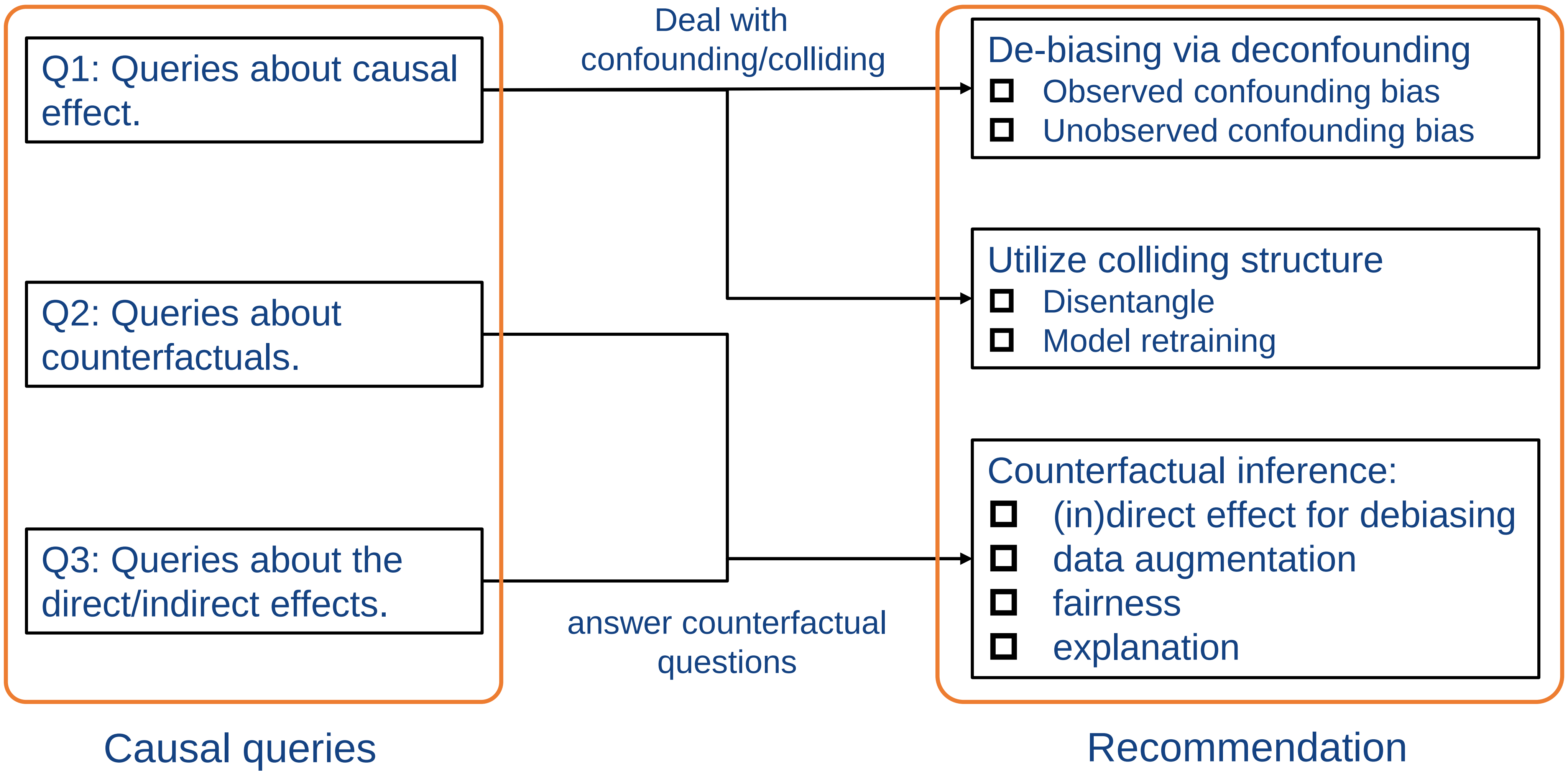
- SCM provides both a mathematical foundation and a friendly calculus for the analysis of causal effects and counterfactuals.
- It can deal with the estimation of three types of causal queries:

  - Queries about the effect of potential interventions.
    To compute causal effect, e.g., $P(Y|do(X))$
  - Queries about counterfactuals.
    e.g., whether event A would occur **if event B had been different**?
  - Queries about the direct / indirect effects. (based on counterfactuals)



the direct effects of X on Y: $X \rightarrow Y$
the indirect effects of X on Y: $X \rightarrow Z \rightarrow Y$

SCM for Recommendation

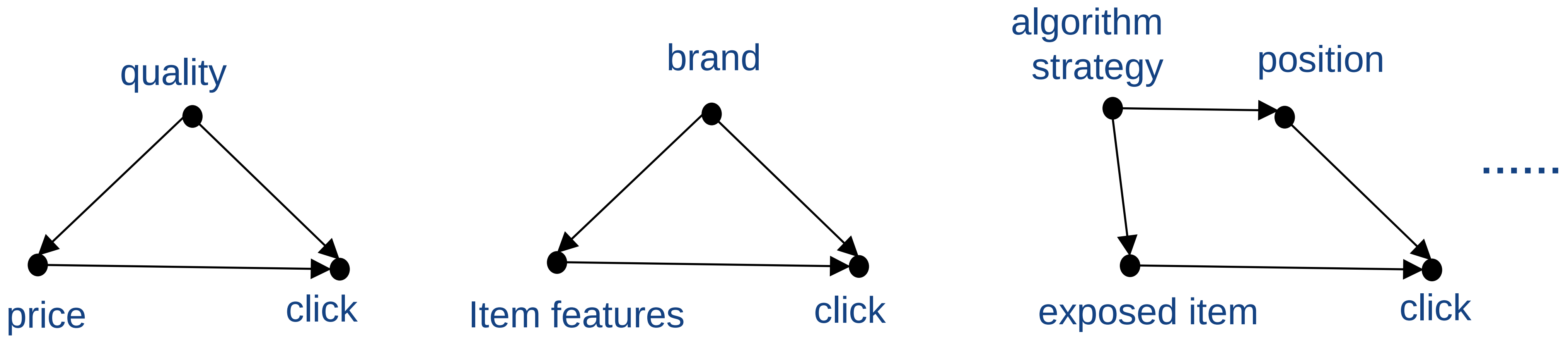


SCM for Recommendation

- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

Confounders in Recommendation

- Are there confounders in recommendation?
  - some examples

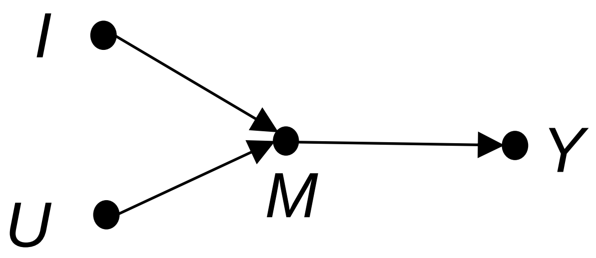


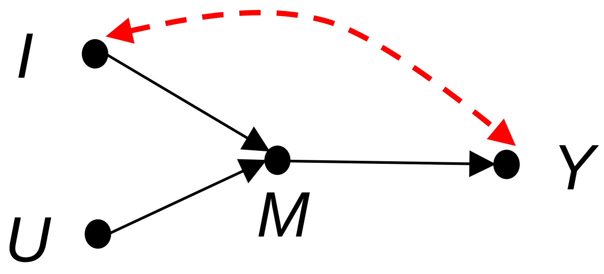
- What’s more, some confounder are observable/measurable, some confounder are unobservable/unmeasurable.

  e.g., company is measurable, quality is unmeasurable.

Confounders in Recommendation

- Is it necessary to deal with confounding effects?
  - The goal of recommendation: estimate user preference. But user preference is implicit.
  - We estimate it as $P(Y|U,I)$, i.e., taking the correlations between (*U*,*I*) pair and click *Y* as the preference.





- However, when there are confounders between U/I and Y(red line), the confounding effect will also bring correlations, while it cannot reflect user preference.

Thus, it is essential to deal with the confounding problem in recommendation!

But HOW?

Existing Work Regarding Observed Confounders

SIGIR TAIPEI | TAIWAN | 2023

The backdoor adjustment is an obvious solution in this line of research.

2021 SIGIR&KDD&CIKM

- Zhang et.al. PDA
- Wang et.al. DecRS
- Yang et.al. DCM
- Gupta et.al. CauSeR

2022TKDE&KDD&CIKM&SIGIR

- Wang et.al. CaDSI
- Zhan et al. D2Q
- He et al. CISD
- Rajanala et al. DeSCoVeR

2023TOIS

- He et.al. DCR
- Zhang et.al. DML

The above work considers different problems caused by confounders, and uses different strategies to implement the backdoor adjustment.

PDA: Confounding View of Popularity Bias

- Popularity bias
  - **Favor a few popular items** while not giving deserved attention to the majority of others
  - The popular items are recommended even more frequently than their popularity would warrant, amplifying long-tail effects.
- Previous methods ignore the underline causal mechanism and blindly remove bias to purchase an even distribution.
- But, not all popularity biases data are bad.
  - Some items have higher popularity because of better quality.
  - Some platforms have the need of introducing desired bias (promoting the items that have the potential to be popular in the future).

Zhang et al. SIGIR 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation

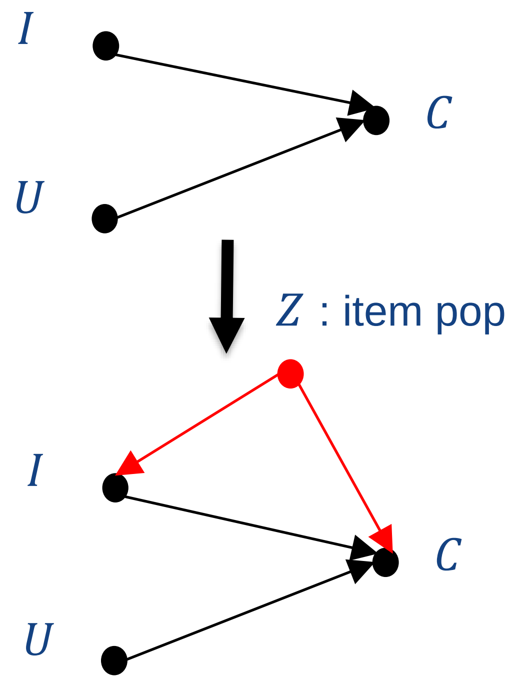
PDA: Confounding View of Popularity Bias

- What is the bad effect of popularity bias?

U: user; **I: exposed item;**
C: interaction label

- Traditional causal assumption
  - $(U, I) \rightarrow C$: user-item matching affects click.
  - Item popularity also has influence on the recommendation process, but is not considered.

- Cofounding view
  - $Z \rightarrow I$: Popularity affects item exposure.
  - $Z \rightarrow C$: Popularity affects click probability.
  - Z is a confounder, bringing spurious (bad effect) correlation between I and C.
  - Take the causation $P(C|do(U, I))$, instead of the correlation $P(C|U, I)$, as user preference.



Causation (backdoor adjustment):
$\boldsymbol{P(C|do(U,I)) = \sum_Z P(C|U,I,Z)P(Z)}$

Correlation:

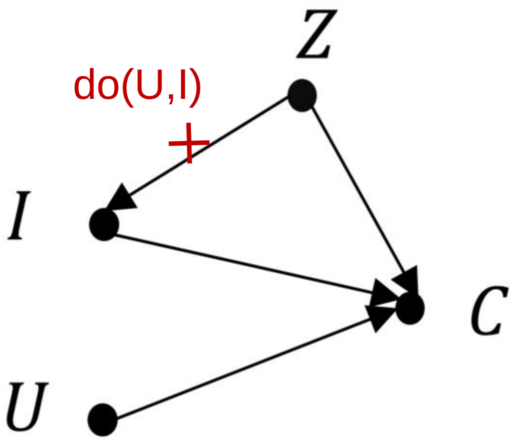
Bad effect

$$P(C|U,I) = \sum_z P(C|U,I,Z)P(Z|I)$$
$$\propto \sum_z P(C|U,I,Z)P(I|Z)P(Z)$$

Zhang et al. SIGIR 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation

PDA: Confounding View of Popularity Bias

- **Training & Inference**: Popularity De-confounding (PD, remove bad effect)



- To estimate $P(C|do(U,I)) = \sum_z P(C|U,I,z)P(z)$
  - **Step 1**. Estimate P(C|U,I,Z)
    - $P_\Theta(c=1|u,i,m_i^t) = f_\Theta(u,i) \times m_i^t$
    - m_i^t the popularity of item i in timestamp t
    - Learn with traditional loss
  - **Step 2**. Compute $P(\mathrm{C}|\mathrm{do}(\mathrm{U},\mathrm{I}))$
    - $\sum_Z P(C|U,I,Z)P(Z) \propto f_\Theta(u,i)$
    - Derivation sees the paper

- **Another Inference**: Popularity Adjusting (inject desired popularity bias)
  - Inject the desired pop bias $\tilde{Z}$ by causal intervention

$$P(C|do(U,I), do(Z=\tilde{z})) \Longrightarrow f_\Theta(u,i) \times \tilde{m}_i$$

Zhang et al. SIGIR 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation

DCR: Deconfounding for Solving Unreliable Label Issue

- **Unreliable label issue:**
  - No ground-truth label for the prediction objective – user preference
  - Only have indirect label: user behaviors

label: user behavior

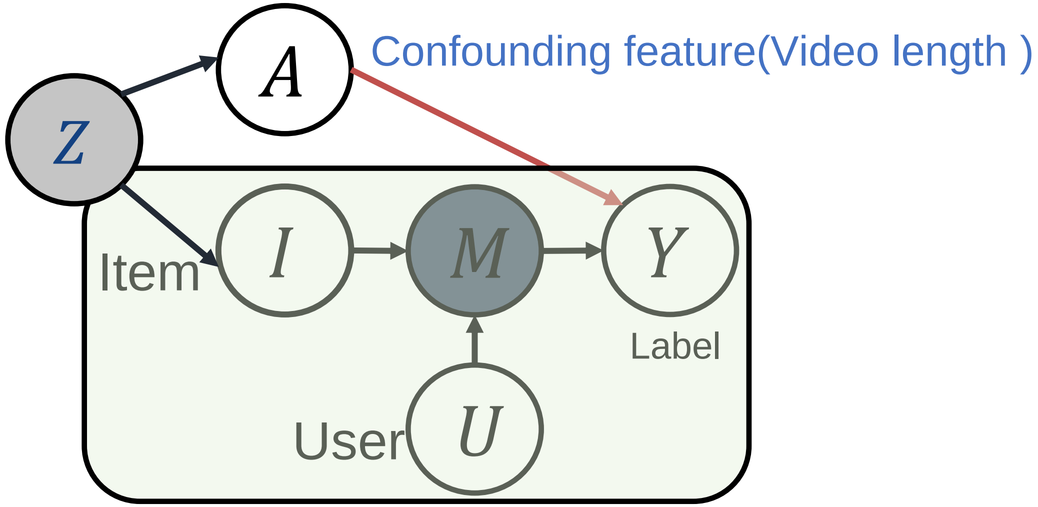
Cannot faithfully reflect

objective: user preference

progress: 60%

progress: 100%

- **Causal Modeling:**
  - Traditional assumption: U-I matching affect label
  - **Some item feature directly affect the label**

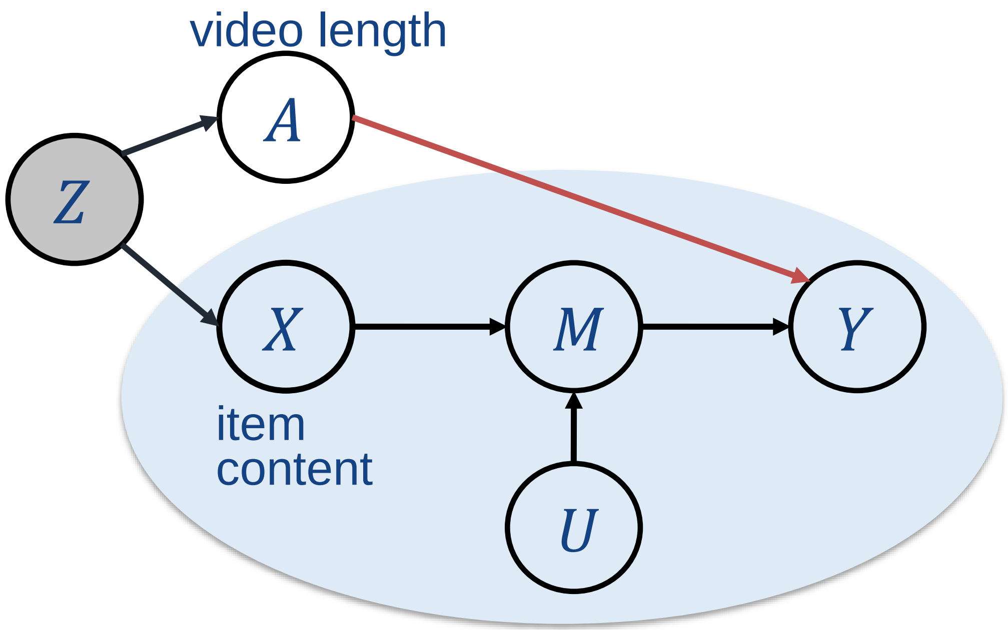


U-I matching (M) partially determines Y

He et al. Addressing Confounding Feature Issue for Causal Recommendation. TOIS 2023.

DCR: Deconfounding for Solving Unreliable Label Issue

Causal analyses



- direct path A→ Y: make P(Y|X, A) biased towards short videos
- Backdoor path $X \leftarrow Z \rightarrow A \rightarrow Y$: make P(Y|X) learn spurious correlation

Should beyond correlation-level

Causal effect as interest

true user preference: the **causal effects** path through M to Y

$$P(Y|U, do(X)) = \sum_{a \in \mathcal{A}} P(Y|U, X, A = a)P(A = a),$$

He et al. Addressing Confounding Feature Issue for Causal Recommendation. TOIS 2023.

DCR: Deconfounding for Solving Unreliable Label Issue

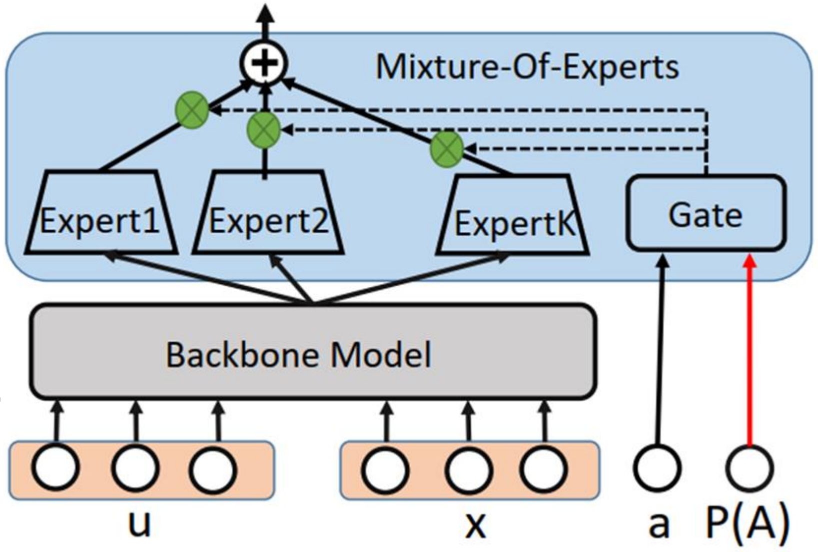
How to estimate the causal effect?

$$P(Y|U, do(X)) = \sum_{a \in \mathcal{A}} P(Y|U, X, A = a) P(A = a),$$

- **DCR: model-based estimation**

$$\boxed{k^{th}\ expert:\ P(y = 1|u, x, A = a_K)}$$

- **Training** --- fitting $P(Y|U, X, A)$
- **Inference** --- backdoor adjustment

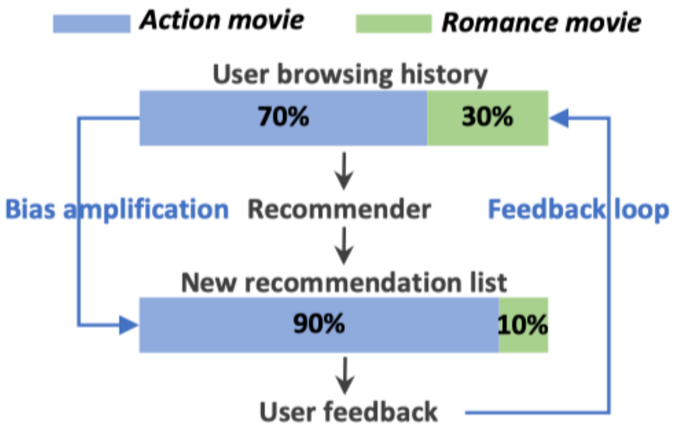


- **DCR involves changing the model architecture, DML [2] proposes to achieve the adjustment directly at the label level/**

【1】He et al. Addressing Confounding Feature Issue for Causal Recommendation. TOIS 2023.
【2】Zhang et al. Leveraging Watch-time Feedback for Short-Video Recommendations: A Causal Labeling Framework. ArXiv 2023.

DecRS: Alleviating Bias Amplification

- **Bias amplification:**
  - What is it?

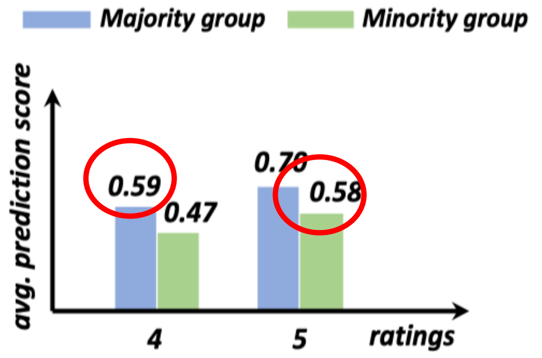


(a) An example of bias amplification.

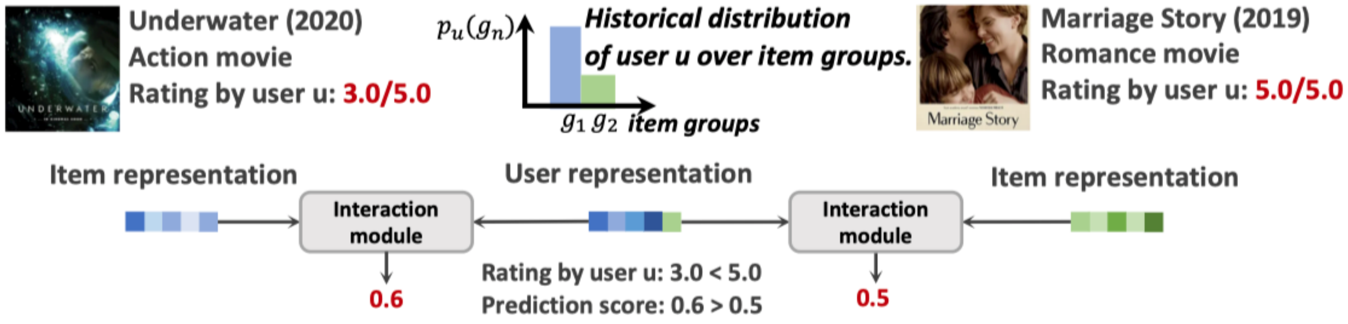
Over-recommend items in the majority group

- **Why?**

- An item with low rating receives a higher prediction score because it belongs to the majority group.
- Intuitively, we can know that the user representation shows stronger preference to majority group.



(b) Prediction score difference between the items in the majority and minority groups over ML-1M.

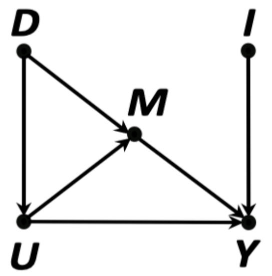


(c) An example on the cause of bias amplification.

Wang et al. SIGKDD 2021. Deconfounded recommendation for alleviating bias amplification.

DecRS: Alleviating Bias Amplification

- **Causal view of bias amplification**



(a)

U	*User representation*
I	*Item representation*
D	*User historical distribution over item groups*
M	*Group-level user representation*
Y	*Prediction score*

- D: user historical distribution over **item group.** $d_u = [p_u(g_1), \ldots, p_u(g_N)]$, e.g., $d_u = [0.8, 0.2]$.
- M: describe how much **the user likes different item groups**, decided by D and U.
- $(U, M) \rightarrow Y$: an item i can have a high Y because: 1) **user's pure preference over the item** ($U \rightarrow Y$) or 2) the **user shows interest in the item group** ($U \rightarrow M \rightarrow Y$).

✓ D is a confounder between U and Y, bringing spurious correlations: given the item i in a group g, the more superior g is in u's history, the higher the prediction score Y becomes.

- **Backdoor adjustment**

$$
\begin{aligned}
&P(Y|U=\mathbf{u}, I=\mathbf{i}) \\
&= \frac{\sum_{\mathbf{d}\in\mathcal{D}}\sum_{\mathbf{m}\in\mathcal{M}} P(\mathbf{d})P(\mathbf{u}|\mathbf{d})P(\mathbf{m}|\mathbf{d},\mathbf{u})P(\mathbf{i})P(Y|\mathbf{u},\mathbf{i},\mathbf{m})}{P(\mathbf{u})P(\mathbf{i})} && \text{(1a)} \\
&= \sum_{\mathbf{d}\in\mathcal{D}}\sum_{\mathbf{m}\in\mathcal{M}} P(\mathbf{d}|\mathbf{u})P(\mathbf{m}|\mathbf{d},\mathbf{u})P(Y|\mathbf{u},\mathbf{i},\mathbf{m}) && \text{(1b)} \\
&= \sum_{\mathbf{d}\in\mathcal{D}} P(\mathbf{d}|\mathbf{u})P(Y|\mathbf{u},\mathbf{i},M(\mathbf{d},\mathbf{u})) && \text{(1c)} \\
&= P(\mathbf{d}_u|\mathbf{u})P(Y|\mathbf{u},\mathbf{i},M(\mathbf{d}_u,\mathbf{u})), && \text{(1d)}
\end{aligned}
$$

➡

$$
\begin{aligned}
&P(Y|do(U=\mathbf{u}), I=\mathbf{i}) \\
&= \sum_{\mathbf{d}\in\mathcal{D}} P(\mathbf{d}|do(U=\mathbf{u}))P(Y|do(U=\mathbf{u}),\mathbf{i},M(\mathbf{d},do(U=\mathbf{u}))) && \text{(2a)} \\
&= \sum_{\mathbf{d}\in\mathcal{D}} P(\mathbf{d})P(Y|do(U=\mathbf{u}),\mathbf{i},M(\mathbf{d},do(U=\mathbf{u}))) && \text{(2b)} \\
&= \sum_{\mathbf{d}\in\mathcal{D}} \boxed{P(\mathbf{d})}P(Y|\mathbf{u},\mathbf{i},M(\mathbf{d},\mathbf{u})), && \text{(2c)}
\end{aligned}
$$

Wang et al. SIGKDD 2021. Deconfounded recommendation for alleviating bias amplification.

DecRS: Alleviating Bias Amplification

- Deconfounded Recommender System (DecRS)
  - To implement:

$$P(Y|do(U=\boldsymbol{u}), I=\boldsymbol{i}) = \sum_{\boldsymbol{d}\in\mathcal{D}} P(\boldsymbol{d})P(Y|\boldsymbol{u},\boldsymbol{i},M(\boldsymbol{d},\boldsymbol{u})) \quad (3)$$

Challenge: the sample space of D is infinite.

- Backdoor adjustment approximation:

(1) Sampling distributions to represent $\mathcal{D}$;

Use function $f(\cdot)$ (FM) to calculate $\mathrm{P}(Y|u,i,M(d,u))$.

$$P(Y|do(U=\boldsymbol{u}), I=\boldsymbol{i}) \approx \sum_{\boldsymbol{d}\in\tilde{\mathcal{D}}} P(\boldsymbol{d})P(Y|\boldsymbol{u},\boldsymbol{i},M(\boldsymbol{d},\boldsymbol{u}))$$
$$= \sum_{\boldsymbol{d}\in\tilde{\mathcal{D}}} P(d)f(\boldsymbol{u},\boldsymbol{i},M(\boldsymbol{d},\boldsymbol{u})) \quad (4)$$

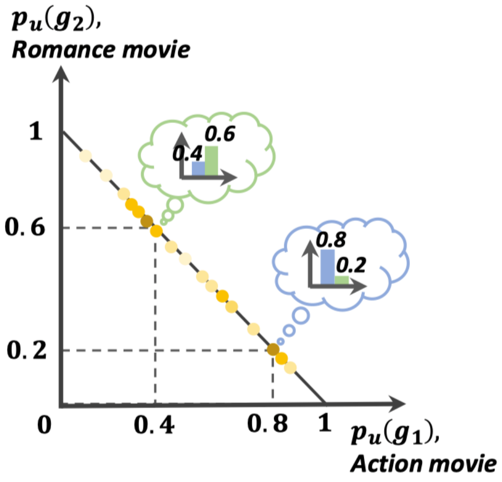
(2) Approximation of $E_d[f(\cdot)]$.

- Expectation of function $f(\cdot)$ of $\boldsymbol{d}$ in Eq. 4 is hard to compute because we need to calculate the results of $f(\cdot)$ for each $\boldsymbol{d}$.
- **Jensen's inequality**: take the sum into the function $f(\cdot)$.

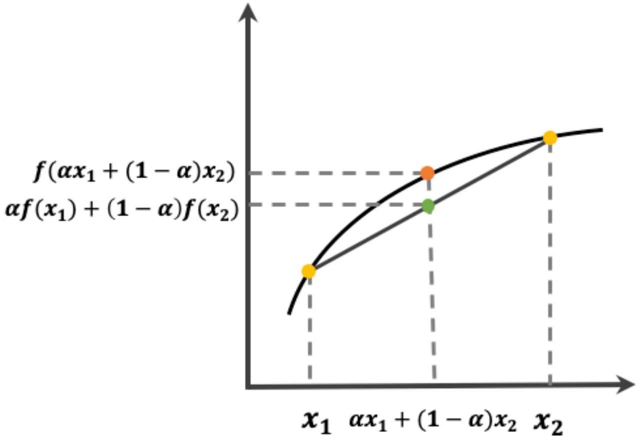
$$P(Y|do(U=\boldsymbol{u}), I=\boldsymbol{i}) \approx \boxed{f(\boldsymbol{u},\boldsymbol{i},M(\sum_{\boldsymbol{d}\in\tilde{\mathcal{D}}} P(\boldsymbol{d})\boldsymbol{d},\boldsymbol{u})).} \quad (5)$$

learn it from data

Different to PDA, this term directly represents the target casual effect.



Infinite Sample Space



Approximation

Wang et al. SIGKDD 2021. Deconfounded recommendation for alleviating bias amplification.

Existing Works for Unobserved Confounders

- The methods based on backdoor adjustment need the confounders could be observable and controllable.
- However, unobserved/unmeasurable/uncontrollable confounders exist in recommendation. How to deal with them?

  - There are two lines of work:

Front-door adjustment

2023 TKDE & TORS — Zhu et.al. CausalD, Xu et al. DCCF

2022 ArXiv — Zhu et.al. HCR

Learning substitutes

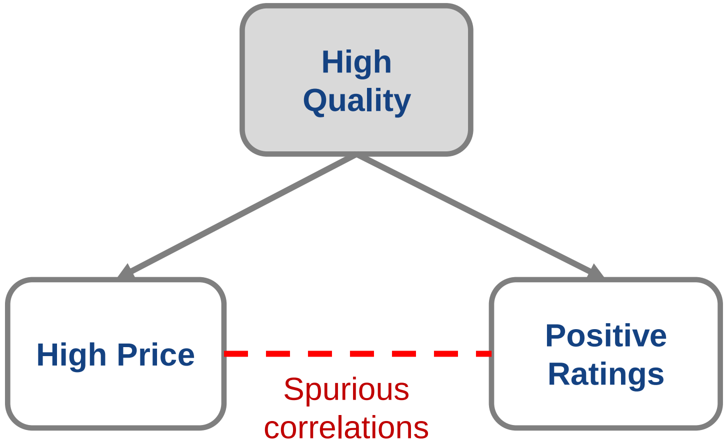
2023 KDD — Zhang et.al. iDCF

2022 ArXiv — Zhu et.al. Deep-Deconf

2020 NeurIPS, 2020 RecSys — Wang et.al. DCF, Zhou et.al. VSR

HCR: Front-door Adjustment-based Solution

- Source of confounding bias is the **confounder** that **affects item attributes and user feedback** simultaneously.
- Some confounders are hard to measure.
  - Technical difficulties, privacy restrictions, *etc*.
  - E.g., product quality.
- Removing hidden confounders is hard:
  - Inverse Propensity Weighting
    - Based on strict assumption of no hidden confounder.
  - Backdoor Adjustment
    - Require the confounder's distribution.

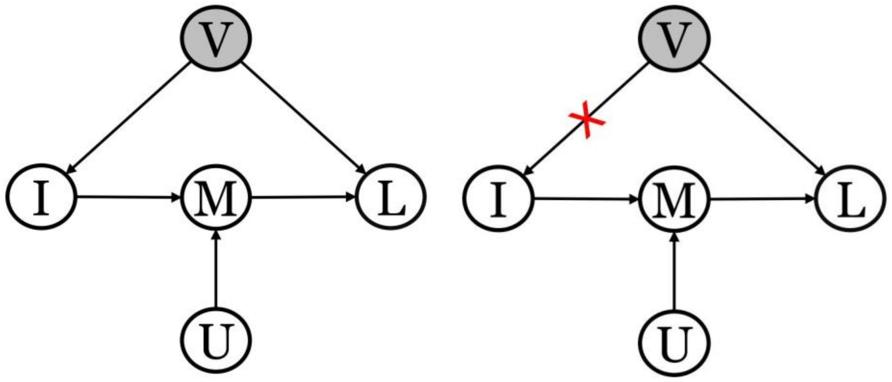


Xinyuan Zhu et.al. "*Mitigating Hidden Confounding Effects for Causal Recommendation*" in 2022.

HCR: Front-door Adjustment-based Solution

- Abstract user feedback generation process into causal graph.
  - V: hidden confounder; L: like feedback; I: item; U: user.
  - M: a set of variables that act as mediators between $\{U, I\}$ and L, e.g., user-item feature matching, and click.
- Key:
  - Block the backdoor path $I \leftarrow V \rightarrow L$
  - Estimate the causal effect of I on L, *i.e.*,

    $P(L|U, \text{do}(I))$.



- Hidden Confounder Removal (HCR) framework.
  - Front-door adjustment
    - decompose causal effect of I on L into: 1) the effects of I on M and 2) the effect of M on L.

$$P\,(L|U, \text{do}\,(I)) = \sum_M P\,(M|U, do(I))P(L|U, do(M))$$
$$= \sum_M P\,(M|U, I) \sum_{I'} P\,(I')P\,(L|M, U, I')$$

Xinyuan Zhu et.al. "*Mitigating Hidden Confounding Effects for Causal Recommendation*" in 2022.

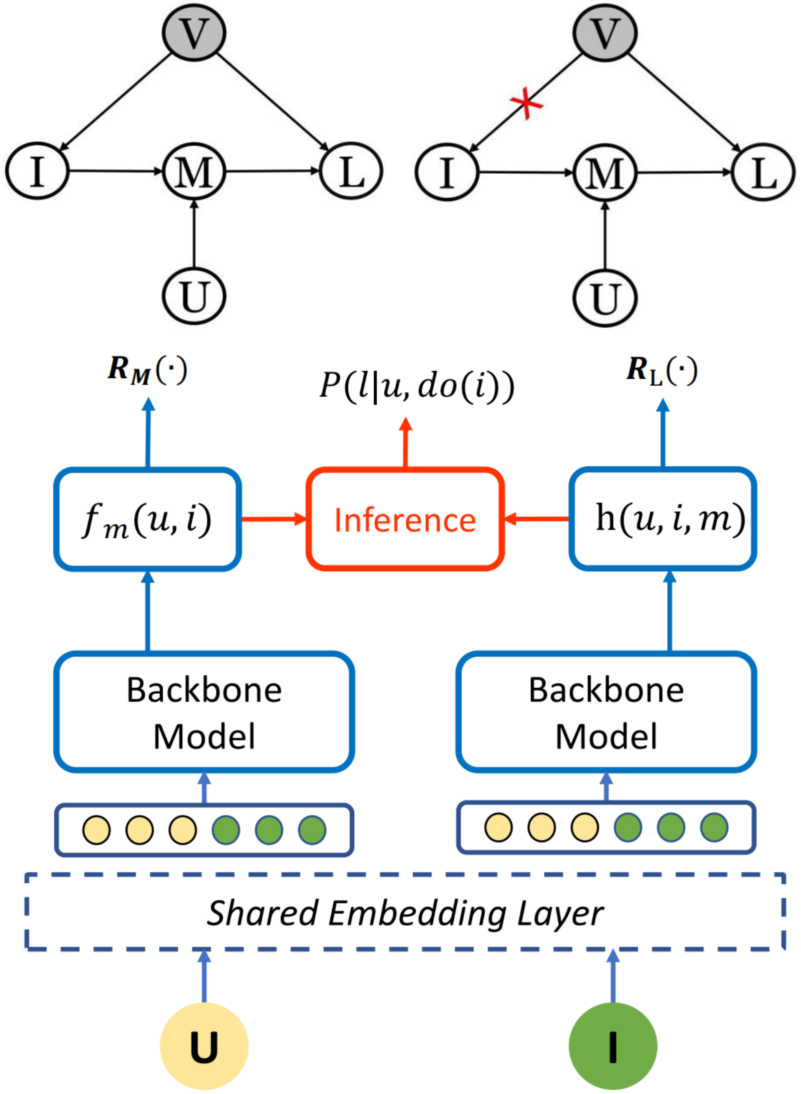
HCR: Front-door Adjustment-based Solution

- Hidden Confounder Removal (HCR) framework
  - $P(L|do(I),U) = \sum_M P\,(M|U,I) \sum_{I'} P\,(I')P\,(L|U,I',M)$
  - Multi-task learning
    - Learns $P(M|U,I) \coloneqq f_m(U,I)$
    - Learn

$$\mathrm{P}(L|M,U,I) \coloneqq h(U,I,M) = h^1(U,M)h^2(U,I')$$

  - Inference
    - Infer $P(M|U,I)$ and $P(L|U,I,M)$
    - Get rid of the sum over I and obtain

$$P(L|U,do(I)) = \sum_M f_m(U,I) \sum_{I'} P(I')h^1(U,M)h^2(U,I') = \sum_M f_m(U,I) h^1(U,M) \sum_{I'} P(I')h^2(U,I') = S_u \sum_M f_m(U,I)h^1(U,M)$$



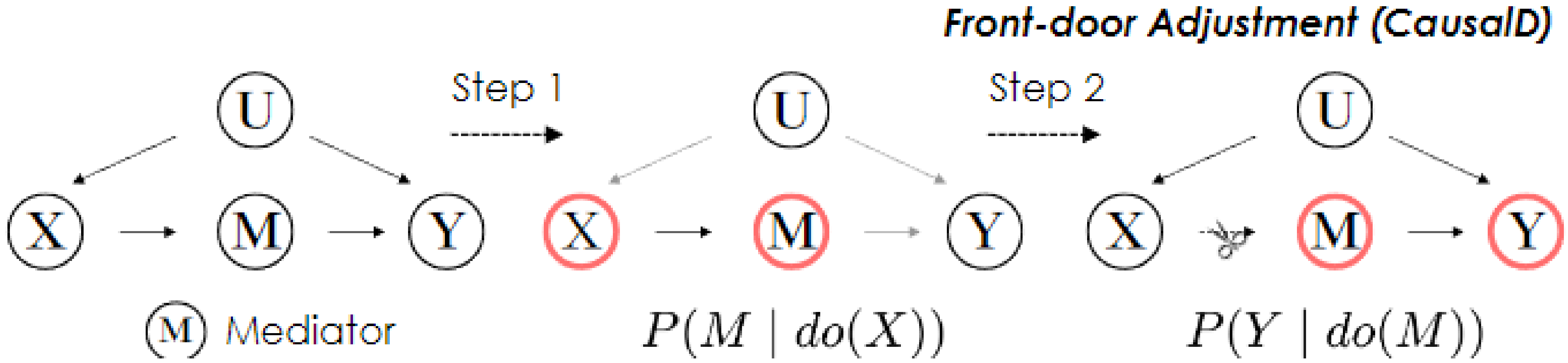
Xinyuan Zhu et.al. "*Mitigating Hidden Confounding Effects for Causal Recommendation*" in 2022.

CausalD: Front-door Adjustment-based Solution

- Consider Hidden Confounder in Sequential Recommendation

Sequential recommendation: predict user next behavior using historical behaviors

X: historical interaction Y: Next behavior M: Representations
U: unobserved confounder, such as social relationships



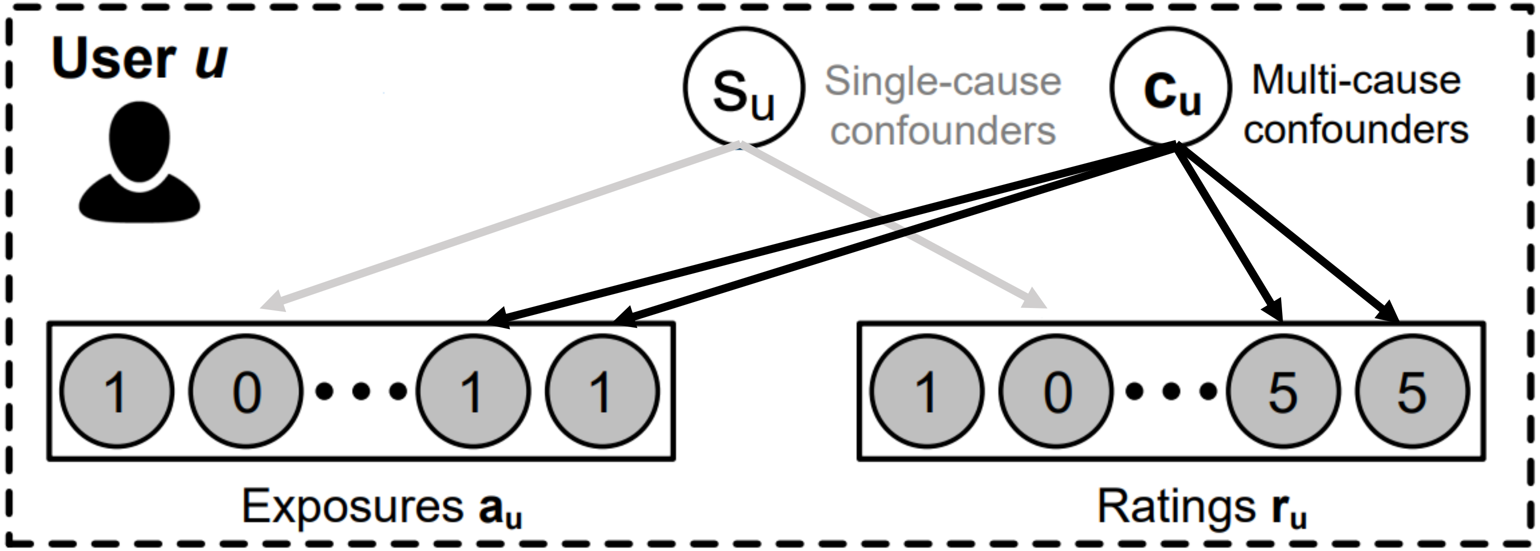
$$P(Y|\mathrm{do}(X)) = \sum_m P(m|do(X))P(Y|do(m))$$
$$= \sum_m P(m|X) \sum_{x'} P(X = x')P(Y|m, x')$$

- Estimation method: similar to HCR but in a distillation manner

Zhang et.al. "Causal Distillation for Alleviating Performance Heterogeneity in recommender System" in TKDE 2023.

Learning Substitutes-based Solution

- Multiple causes assumption for recommendation:
  - multiple causes: each user's binary exposure to an item a_{ui} is a cause(treatment), thus there are multiple causes.
  - There are multiple-cause confounders (confounders that affect ratings and many causes).
  - Single-cause confounders (confounders that affect ratings and only one cause) are negligible.



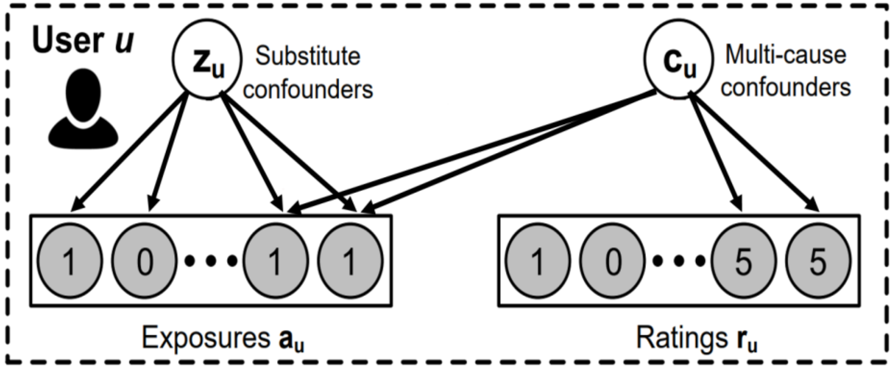
Wang et al. RecSys 2020. Causal inference for recommender system.
Zhu et.al. Arxiv 2022. Deep causal reasoning for recommendations.

Learning Substitutes-based Solution

- Learning substitutes to deconfounding:

Key: if Z_u renders the $a_{u,i}$'s conditionally independent then there cannot be another multi-cause confounder

Contradiction: assume $p(a_{u1}, \dots, a_{um}|z_u) = \prod_i p(a_{ui}|z_u)$, if there is a multi-cause confounder, the conditional independence cannot hold.



- Step 1: learning substitutes

Finding a Z_u, such that:

$$p(a_{u1}, \dots, a_{um}|z_u) = \prod_i p(a_{ui}|z_u)$$

Example:

find a generative model:

$$P_\Theta(A_u|Z_u) = \prod_{i=1}^{m} Bern(a_{ui}|\theta(z_u)_i)$$

then:

find $q_\Phi(Z_u|A_u)$ with variation-inference

- Step 2: deconfounded recommender

Control the substitutes to fit recommender model

Example:

$$y_{ui}(a) = \theta_u^\top \beta^i \cdot a + \gamma_u \cdot z_{ui} + \epsilon_{ui}$$

where θ_u and β_i refer user preference and item attributes, respectively.

Wang et al. RecSys 2020. Causal inference for recommender system.
Wang et.al. *J Am Stat Assoc 2019.* The blessings of multiple causes.
Zhu et.al. Arxiv 2022. Deep causal reasoning for recommendations.

Papers on Deconfounding Recommendation

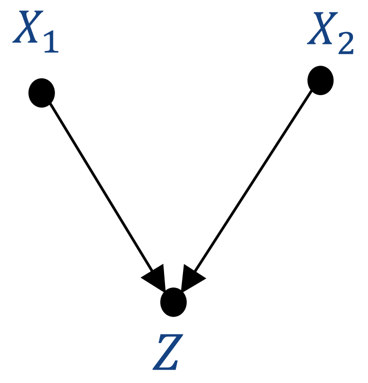
- Zhang, Yang, et al. "Causal intervention for leveraging popularity bias in recommendation." In *SIGIR* 2021. (Zhang et.al. PDA)
- Wang, Wenjie, et al. "Deconfounded recommendation for alleviating bias amplification." In *SIGKDD* 2021. (wang et.al. DecSR)
- Wang, Xiangmeng, et al. "Causal Disentanglement for Semantics-Aware Intent Learning in Recommendation." In *TKDE* 2022. (Wang et.al. CaDSI)
- Gupta, Priyanka, et al. "CauSeR: Causal Session-based Recommendations for Handling Popularity Bias." In *CIKM* 2021. (Gupta et.al., CauSeR)
- Rajanala, Sailaja, et al. "Descover: Debiased semantic context prior for venue recommendation.“ In *SIGIR* 2022 (Rajanala et al. DeSCoVeR)
- Yang, Xun, et al. "Deconfounded video moment retrieval with causal intervention." In *SIGIR 2021*. (Yang et.al. DCM)
- Zhan, Ruohan, et al. "Deconfounding Duration Bias in Watch-time Prediction for Video Recommendation.“ SIGKDD 2022. (Zhan et al. D2Q)
- He, Ming, et al. "Causal intervention for sentiment de-biasing in recommendation." *In CIKM 2022*. (He et al. CISD)
- He, Xiangnan, et al. "Addressing confounding feature issue for causal recommendation." *ACM TOIS* 2023. (He et al. DCR)
- Wang, Yixin, et al. "Causal inference for recommender systems." Fourteenth ACM Conference on Recommender Systems. 2020. (Wang et.al. DCF)
- Zhang, Yang, et al. "Leveraging Watch-time Feedback for Short-Video Recommendations: A Causal Labeling Framework." arXiv 2023. (Zhang et al. DML)
- S. Zhang *et al*., "Causal Distillation for Alleviating Performance Heterogeneity in Recommender Systems,“ TKDE 2023. (Zhang et al. CausalD)
- Qing Zhang et.al. Debiasing Recommendation by Learning Identifiable Latent Confounders. KDD 2023. (Zhang et al. iDCF)
- Zhu, Xinyuan, et al. "Mitigating hidden confounding effects for causal recommendation." *arXiv 2022. (Zhu et al. HCR)*

SCM for Recommendation

- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

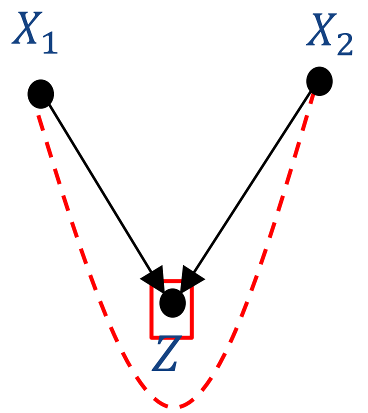
Colliding Effects in Recommendation

- Are there colliders in recommendation?
  - There are variables affected by many factors. Such as, the happening of clicking is affected by user preference and the exposure position.
  - Existing work also tries to construct colliders manually.



- To utilize or eliminate colliding effects?
  - Assume that we have known X_2, try to estimate X_1.
  - Condition on Z, X_1 and X_2 could be correlated.
  - That means condition on Z, X_2 would provide us more information to estimate X_1.

In recommendation, we usually face with this case (know X_2 and Z to predict X_1). Thus existing work based on SCM tries to utilize colliding effects to better learn some targets.



DICE: Colliding Effects for Disentangling True Interest

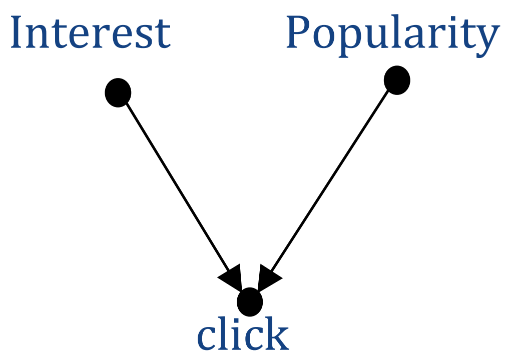
- What are causes of a user-item interaction (click)?

Two main causes:
- Interest
- Conformity

User tend to follow the mainstream

- Disentangle Interest and Conformity to identify true interest.
- But it is hard because of lacking ground-truth. (An interaction can come from either factor or both factors)

- Colliding effect can come to help:



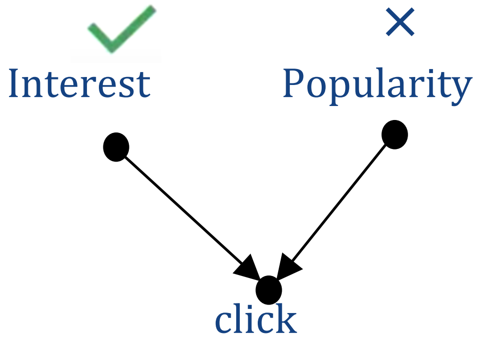
- Interest and Popularity (conformity) are independent
- But, they are correlated given clicks:

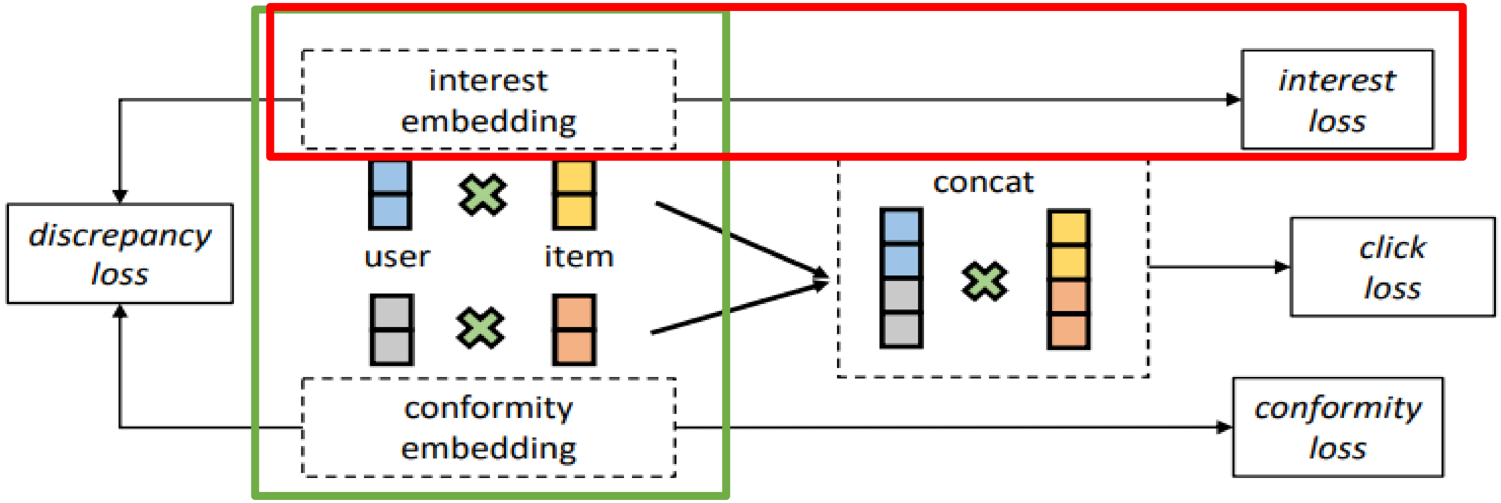
A click on less popular item → High Interest

Zheng et al. WWW 2021. Disentangling User Interest and Conformity for Recommendation with Causal Embedding

DICE: Colliding Effects for Disentangling True Interest

- DICE: Partial pairwise data identifies true interest:
  - O_1: {<u, pos_item, neg_item>, wherein *pos_item* is less popular than *neg_item*}
  - Pairwise cause-specific data (interst-driven): we can ascertain that the interaction is more likely due to user interest





- **Key 2: learning interest embedding on interest-driven pairwise data (O_1).**

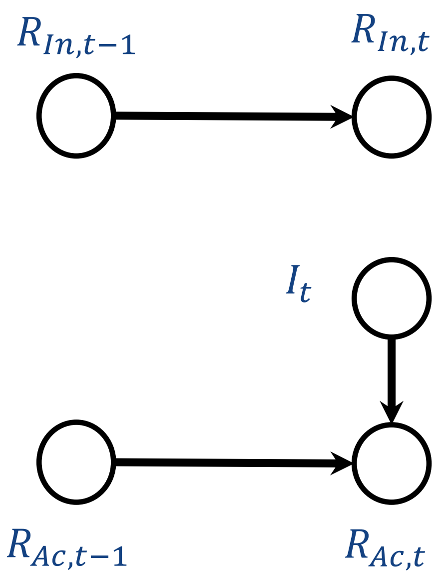
- **Key1: split user/item representation into two embeddings**

- The core idea of leveraging colliding effects has also been extended to Sequential Recommendation. (Sun et al. MiceRec. 2022.)

Zheng et al. Disentangling User Interest and Conformity for Recommendation with Causal Embedding. In WWW 2021.
Sun et al. Multi-interest Sequence Modeling for Recommendation with Causal Embedding. In SDM 2022. (Sun et al. MiceRec.)

Colliding Effects for Incremental Training

- Incremental training for recommender system
  - Usually, using the incremental interaction data I_t for efficient retraining.
  - Only updating the representations of **active** user/item corresponding to I_t.
  - Ignoring the representations of **inactive** user/item.



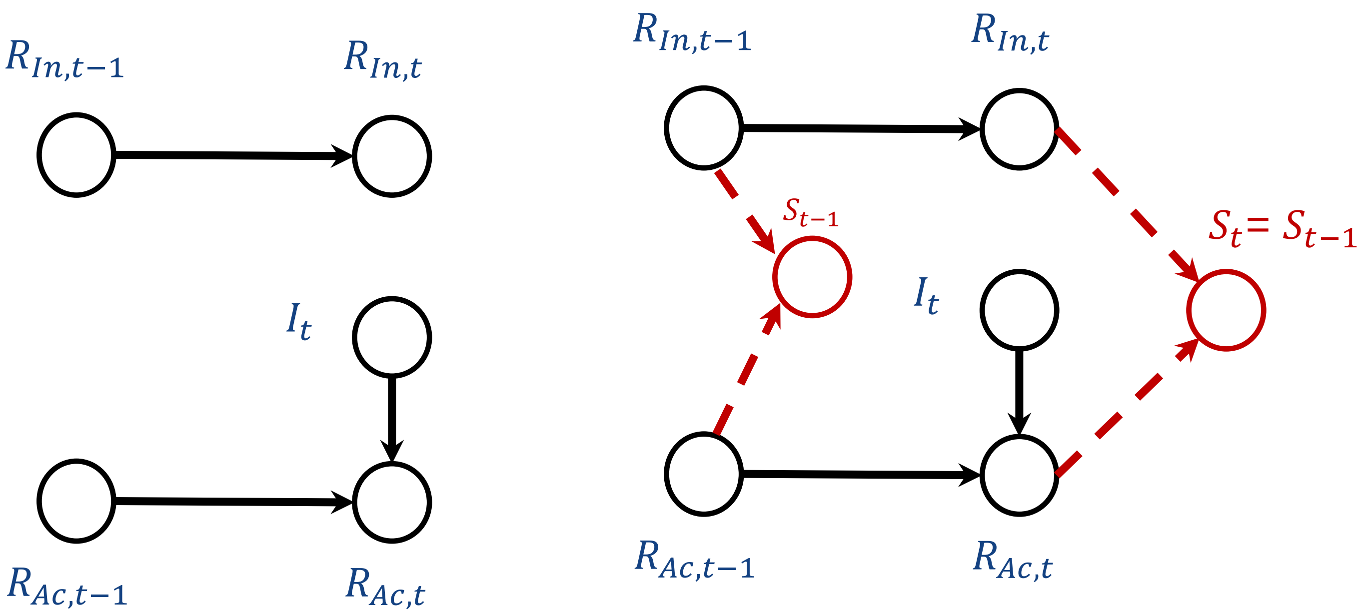
Causal graph of incremental training

- $R_{In,t-1}$: Representations of inactivate user/item at time *t-1*.
- $R_{In,t}$: Representations of inactivate user/item at time *t*.
- $R_{Ac,t-1}$: Representations of activate user/item at time *t-1*.
- $R_{Ac,t}$: Representations of activate user/item at time *t*.
- I_t: Incremental interaction data collected from time *t-1* to *t*.

Ding, Sihao, et al. "Causal incremental graph convolution for recommender system retraining." IEEE TNNLS (2022).

Colliding Effects for Incremental Training

- Causal incremental training with colliding effects



Building colliding effect

- Creating a collider S_t between $R_{In,t}$ and $R_{Ac,t}$, S_t is the similarity between representations of active and inactivate user/item.
- Restraining $S_t = S_{t-1}$ to open the causal path $I_t \rightarrow R_{Ac,t} \rightarrow R_{In,t}$ with the help of colliding effect.
- Using the incremental data I_t simultaneously update both $R_{Ac,t}$ and $R_{In,t}$.

Ding, Sihao, et al. "Causal incremental graph convolution for recommender system retraining." IEEE TNNLS (2022).

SCM for Recommendation

- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

Counterfactual Recommendation

- Counterfactual inference for debiasing
  - Focus on **removing path-specific effects** for debiasing
  - First estimate the causal effect by comparing a counterfactual world with the factual world, and then mitigate path-specific effects.

- Representative Work
  - Wang, et al. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In SIGIR 2021.
  - Wei, et al. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In KDD 2021.
  - Zihao Zhao et al. Popularity Bias Is Not Always Evil: Disentangling Benign and Harmful Bias for Recommendation. In TKDE (2022).
  - Gang Chen et al. Unbiased Knowledge Distillation for Recommendation. In WSDM 2023.

Counterfactual for Mitigating Clickbait Bias

- **Clickbait bias**
  - User interactions are biased to the items with attractive exposure features.
  - Clickbait items: exposure features (*e.g.*, title/cover image) attract users while content features (*e.g.*, video) are disappointing.
  - Recommender models learned from the biased interactions will frequently recommend these clickbait items, decreasing user experience.

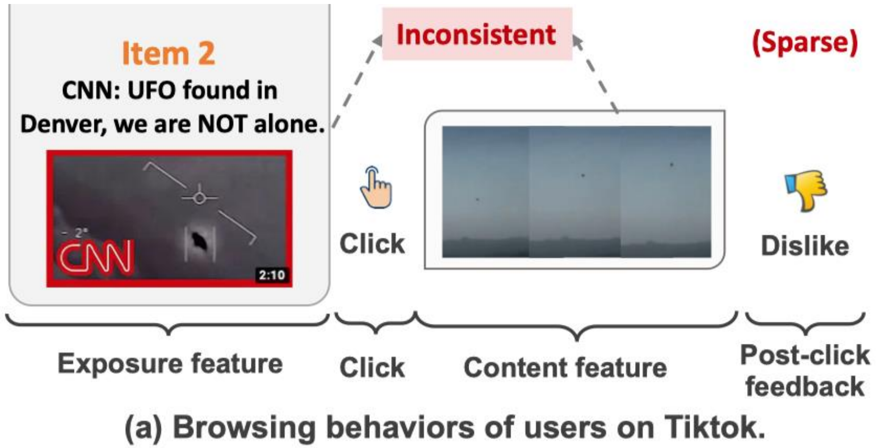
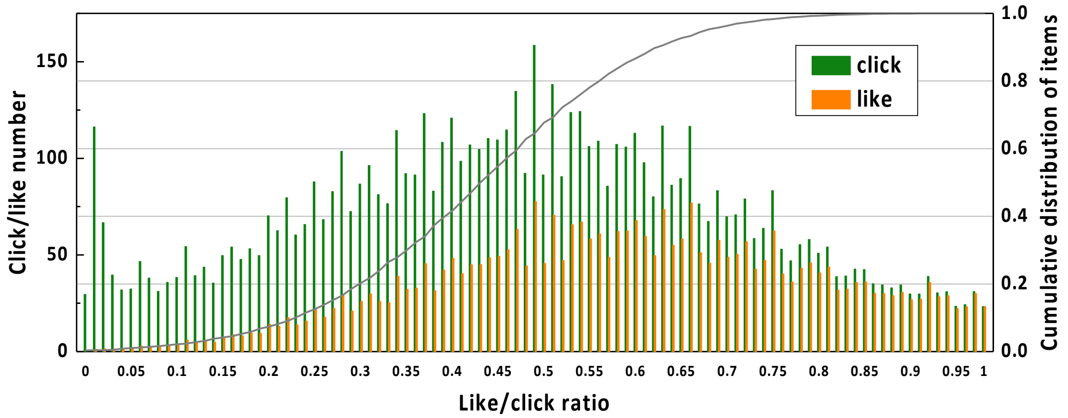


Fig. Statistics of clicks and likes on Tiktok dataset. Partly show the wide existence of clickbait issue.



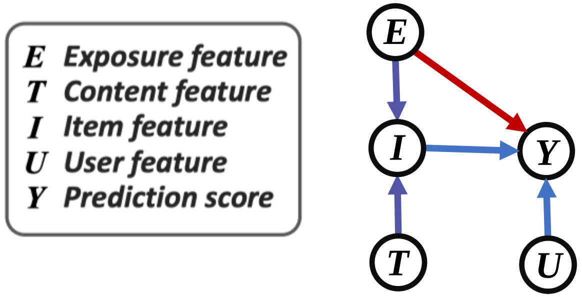
Wang et al. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In SIGIR 2021.

Counterfactual for Mitigating Clickbait Bias

- ## Counterfactual Inference

❖ **Causal Graph**

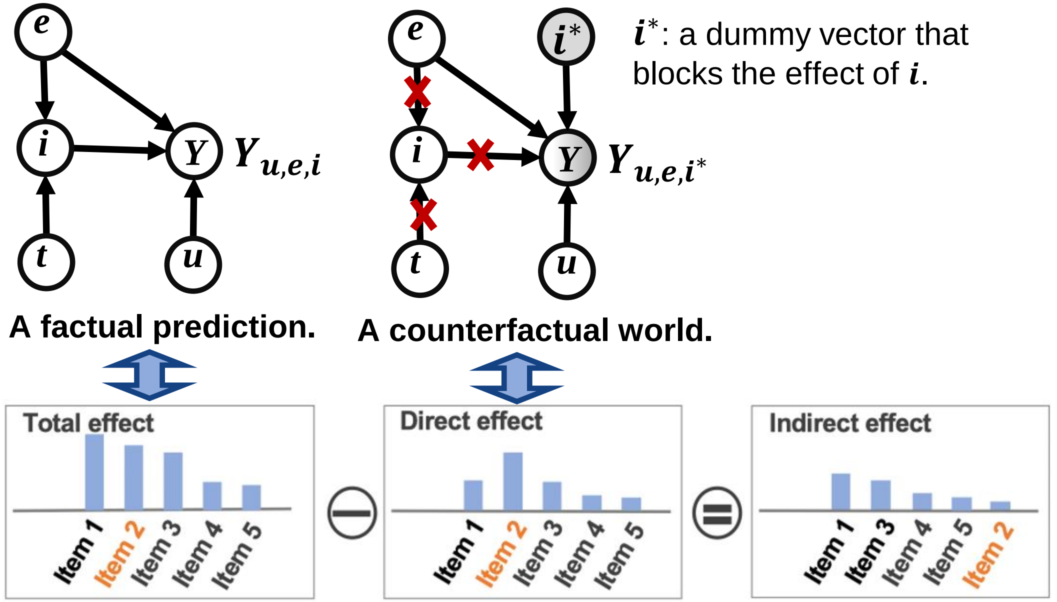
- A causal graph to describe the causal relationships between the features and user-item prediction scores.
- **Reason for clickbait issue:** $E \rightarrow Y$ a clickbait item has high prediction scores purely due to its attractive exposure features, *i.e.*, title/cover.



❖ **Causal learning** for training**:** learn structural functions $I(E,T)$ and $Y(U,I,E)$ from data.

❖ **Causal reasoning** for inference**:** counterfactual inference.

- Reduce the direct effect of exposure features.
- **1) Estimate** the effect in the counterfactual world, which imagines *what the prediction score would be if the item had only the exposure features.*
- **2) Reduce** the direct effect of exposure features for inference.



Wang et al. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In SIGIR 2021.

Counterfactual for Mitigating Clickbait Bias

• Overall Performance

Table 2: Top-K recommendation performance of compared methods on Tiktok and Adressa. %Improve. denotes the relative performance improvement of CR over NT. The best results are highlighted in bold. Stars and underlines denote the best results of the baselines with and without using additional post-click feedback during training, respectively.

Dataset	Tiktok						Adressa					
Metric	P@10	R@10	N@10	P@20	R@20	N@20	P@10	R@10	N@10	P@20	R@20	N@20
NT [50]	0.0256	0.0357	0.0333	0.0231	0.0635	0.0430	0.0501	0.0975	0.0817	0.0415	0.1612	0.1059
CFT [50]	0.0253	0.0356	0.0339	0.0226	0.0628	0.0437	0.0482	0.0942	0.0780	0.0405	0.1573	0.1021
IPW [27]	0.0230	0.0334	0.0314	0.0210	0.0582	0.0406	0.0419	0.0804	0.0663	0.0361	0.1378	0.0883
CT [50]	0.0217	0.0295	0.0294	0.0194	0.0520	0.0372	0.0493	0.0951	0.0799	0.0418*	0.1611	0.1051
NR [51]	0.0239	0.0346	0.0329	0.0216	0.0605	0.0424	0.0499	0.0970	0.0814	0.0415	0.1610	0.1058
RR	0.0264*	0.0383*	0.0367*	0.0231*	0.0635*	0.0430*	0.0521*	0.1007*	0.0831*	0.0415	0.1612*	0.1059*
CR	**0.0269**	**0.0393**	**0.0370**	**0.0242**	**0.0683**	**0.0476**	**0.0532**	**0.1045**	**0.0878**	**0.0439**	**0.1712**	**0.1133**
%Improve.	5.08%	10.08%	11.11%	4.76%	7.56%	10.70%	6.19%	7.18%	7.47%	5.78%	6.20%	6.99%

- **Observations:**
  - **RR** achieves the best performance in the baselines by using post-click feedback for reranking.
  - **Proposed CR** significantly recommends more satisfying items by mitigating clickbait bias.

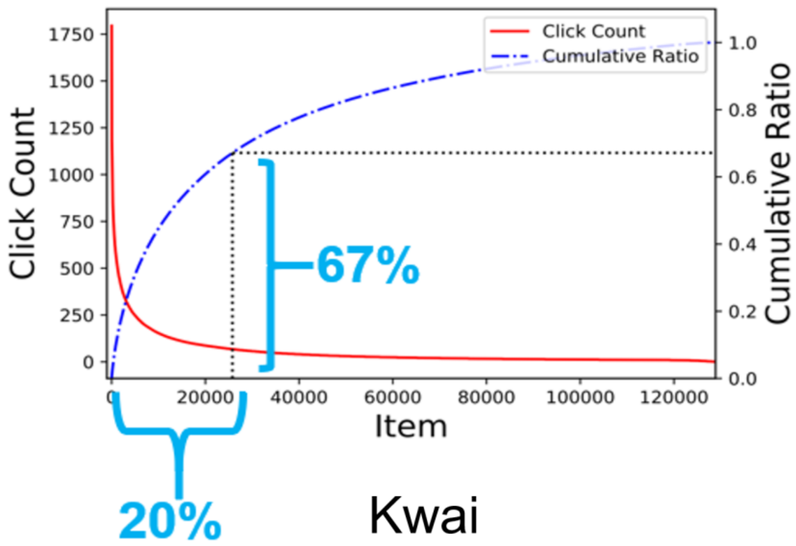
Wang et al. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In SIGIR 2021.

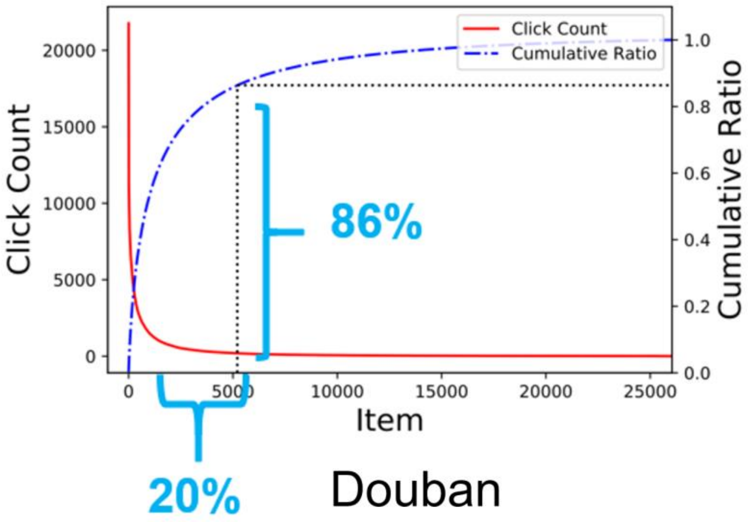
Counterfactual for Mitigating Popularity Bias

Popularity Bias in RecSys

- Popularity bias ≠ Uneven popularity distribution
  - The popular items are gradually over-recommended, amplifying long-tail effects.
  - Favor a few popular items while not giving deserved attention to the majority of others.

- From data perspective:



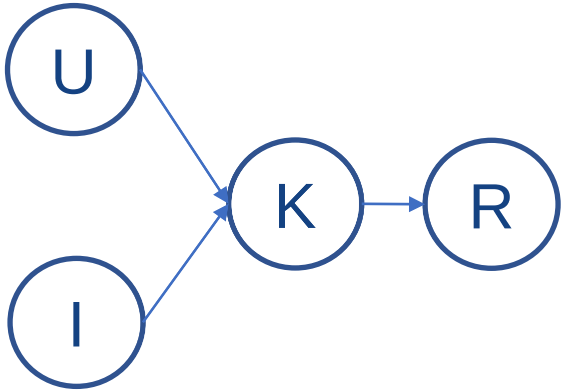


Long-tail distribution

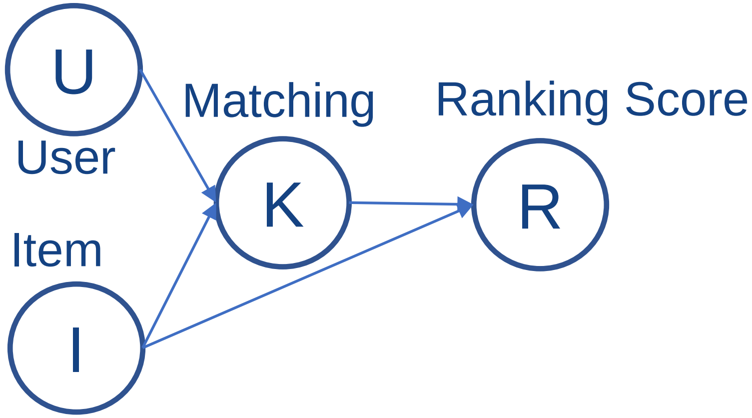
Wei et al. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. In KDD 2021.

Counterfactual for Mitigating Popularity Bias

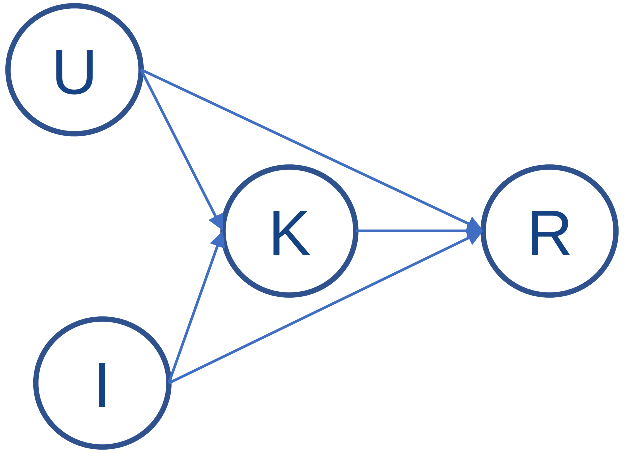
- Causal View of Popularity Bias



Common Recommender
User-Item Matching



Popularity bias modeling:
Incorporating item popularity



User-specific modeling:
Incorporating item popularity & user activity

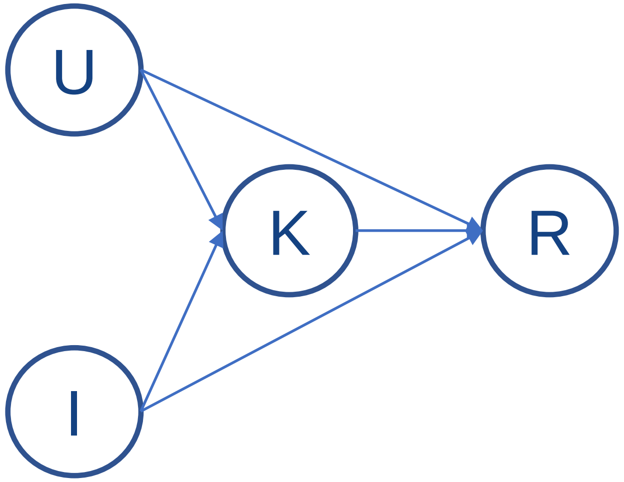
  - Edge I→R captures popularity bias.
  - Edge U→R captures the user sensitive to popularity.

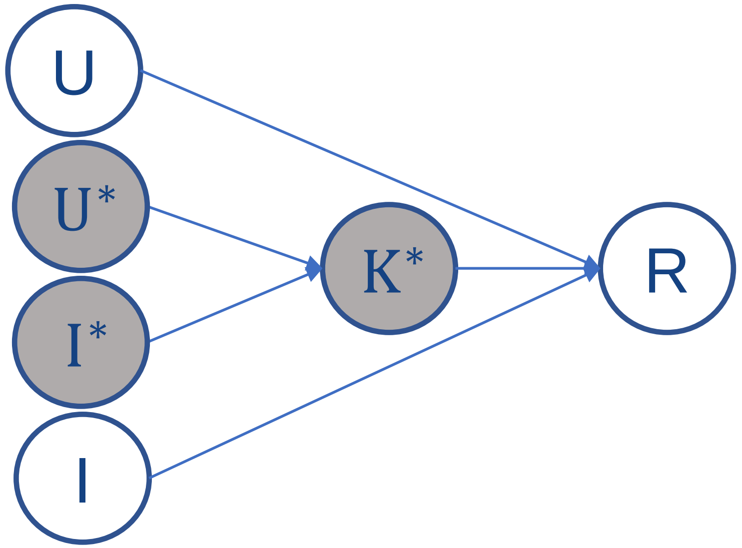
- Solution:
  - Train a recommender based on the causal graph via a multi-task learning
  - Perform counterfactual inference to eliminate popularity bias (*Question to answer: what would the prediction be if there were only popularity bias?*)

Wei et al. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. In KDD 2021.

Counterfactual for Mitigating Popularity Bias

- **Counterfactual Inference to Remove Bias**
- *Question: what the prediction would be if there were no bias?*





$$TIE = TE - NDE = Y(U=u, I=i, K=K_{u,i}) - Y(U=u, I=i, K=K_{u^*,i^*})$$

Factual world Counterfactual world

Inference with TIE $= \hat{y}_k \times \sigma(\hat{y}_i) \times \sigma(\hat{y}_u)$ - $c \times \sigma(\hat{y}_i) \times \sigma(\hat{y}_u)$

Wei et al. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. In KDD 2021.

Counterfactual for Mitigating Popularity Bias

- Evaluate MACR framework on two base models: MF and LightGCN.
- Testing data is intervened to be uniform.

MF as the backbone

Method \ data	Adressa		Yelp2018	
	Recall	NDCG	Recall	NDCG
MF	0.0853	0.0341	0.0060	0.0094
ExpoMF	0.0896	0.0365	0.0060	0.0093
MF_causE	0.0835	0.0365	0.0051	0.0083
MF_BS	0.0900	0.0377	0.0061	0.0098
MF_reg	0.0659	0.0332	0.0050	0.0081
MF_IPS	**0.0964**	**0.0392**	**0.0062**	**0.0100**
MACR	0.1090	0.0495	0.0264	0.0192

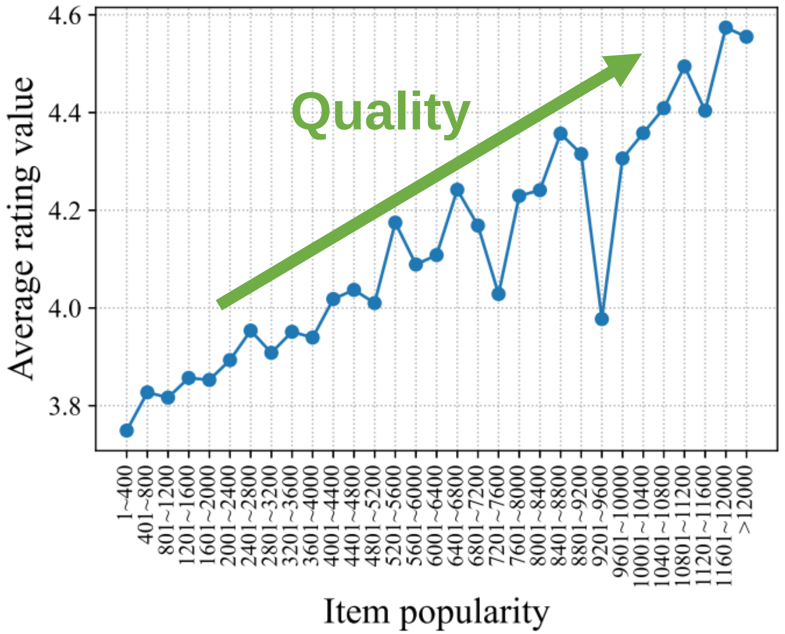
LightGCN as the backbone

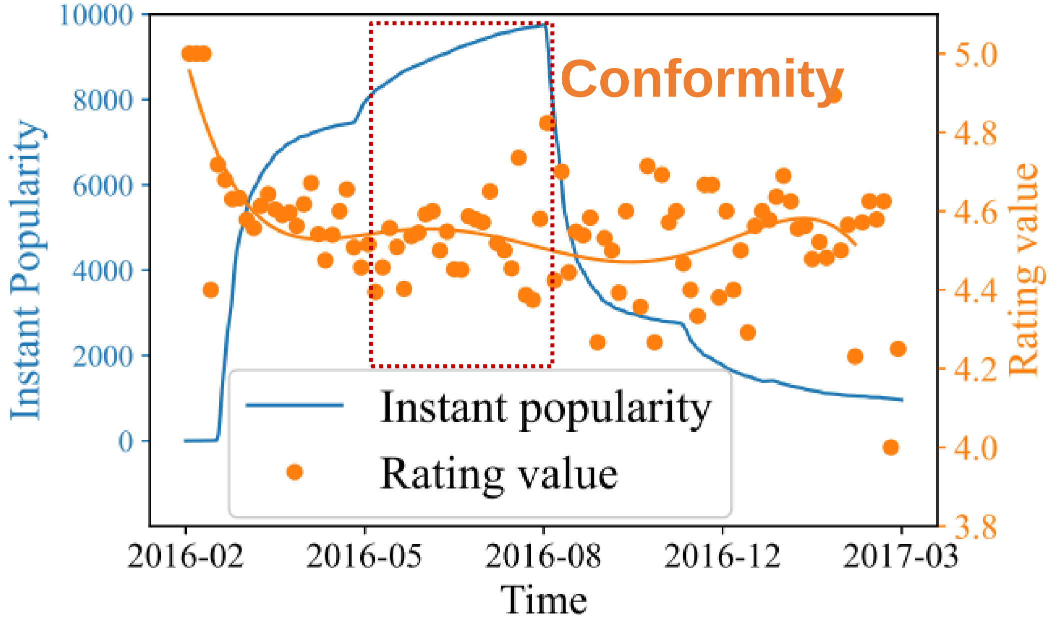
Method \ data	Adressa		Yelp2018	
	Recall	NDCG	Recall	NDCG
Lgcn	0.0977	0.0395	0.0044	0.0086
Lgcn_causE	0.0823	0.0374	0.0050	0.0088
Lgcn_BS	0.1085	0.0469	**0.0048**	**0.0088**
Lgcn_reg	**0.0979**	**0.0390**	0.0042	0.0083
Lgcn_IPS	0.1070	0.0468	0.0054	0.0090
MACR	0.1273	0.0525	0.0312	0.0177

Wei et al. Model-Agnostic Counterfactual Reasoning for Eliminating Popularity Bias in Recommender System. In KDD 2021.

Counterfactual for Leveraging Popularity Bias

- **Conflicting Observation:**
  - The more **popular** an item is, the larger average **rating value** the item tends to have (**positive** correlation).
  - From the temporal view, for a large proportion of items, the **rating value** exhibits **negative** correlation with the item **popularity** at that time
- Quality + Conformity → Popularity, thus disentangle benign and harmful Bias

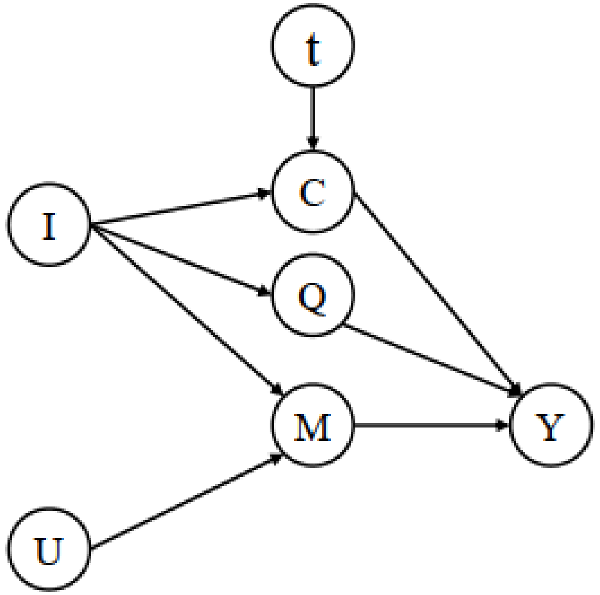




Zhao et al. Popularity Bias is not Always Evil: Disentangling Benign and Harmful Bias for Recommendation. TKDE' 22.

Counterfactual for Leveraging Popularity Bias

- Time-aware **D**is**E**ntangled framework(**TIDE**)
  - Main challenge: Lack of explicit signal for disentanglement
- **Quality** is **static:** $\mathrm{I} \rightarrow Q \rightarrow Y$
  - Quality has **stable** influence on users' behavior
- **Conformity** is **dynamic:** $(\mathrm{I}, \mathrm{t}) \rightarrow C \rightarrow Y$
  - Conformity is **time-sensitive**
- **User interest:** $(\mathrm{U}, \mathrm{I}) \rightarrow M \rightarrow Y$
  - User and item's matching score, can be Implemented by various recommendation models, such as MF, LightGCN, etc.



(a) Causal graph of our TIDE.

U: User I: Item
t: time C: conformity
Q: Quality Y: Prediction
M: Matching score

Zhao et al. Popularity Bias is not Always Evil: Disentangling Benign and Harmful Bias for Recommendation. TKDE' 22.

Counterfactual for Leveraging Popularity Bias

- **Training Stage**:
  - **Popularity** comes from **Quality** and **Conformity**
  - Prediction with **Popularity** and **matching score**

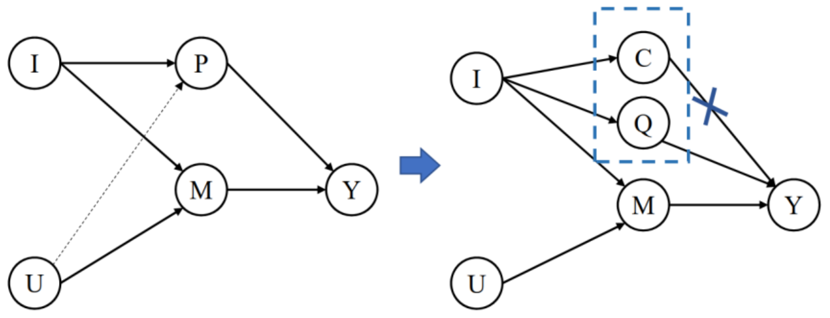
$$\hat{y}_{ui}^{t} = \mathrm{Tanh}(q_i + c_i^t) \times \mathrm{Softplus}(m_{ui})$$

- **Inference Stage**:
  - Intervention: set c as reference vector c^* (e.g., zero) during inference to **remove** the **improper effect from C to Y**.

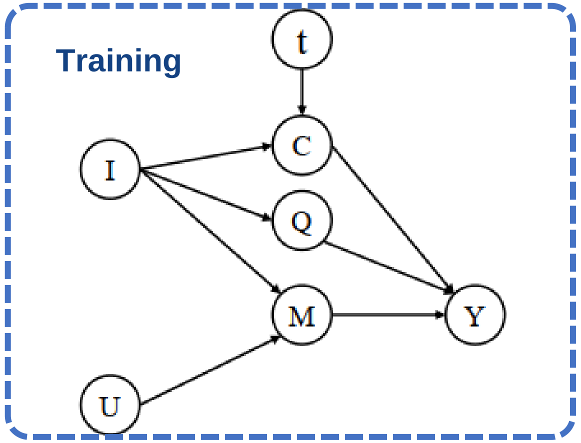
$$\hat{y}_{ui}^{*} = \tanh(q_i + c^*) \times \mathrm{Softplus}(m_{ui})$$

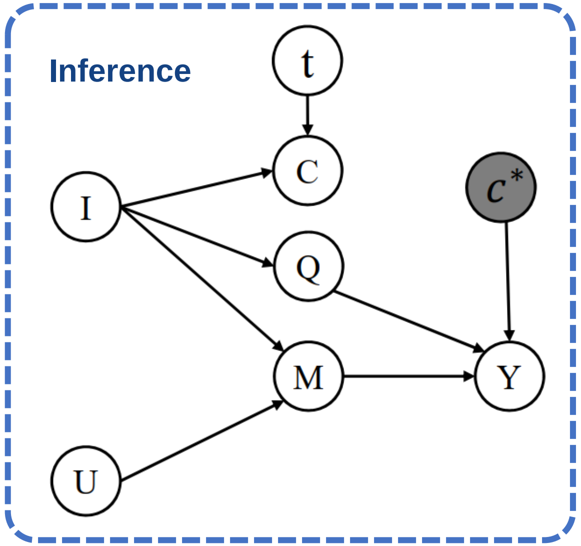
- Comparison with PD
  - TIDE further conduct disentanglement of popularity bias

PD



TIDE





Zhao et al. Popularity Bias is not Always Evil: Disentangling Benign and Harmful Bias for Recommendation. TKDE' 22.

SCM for Recommendation

- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

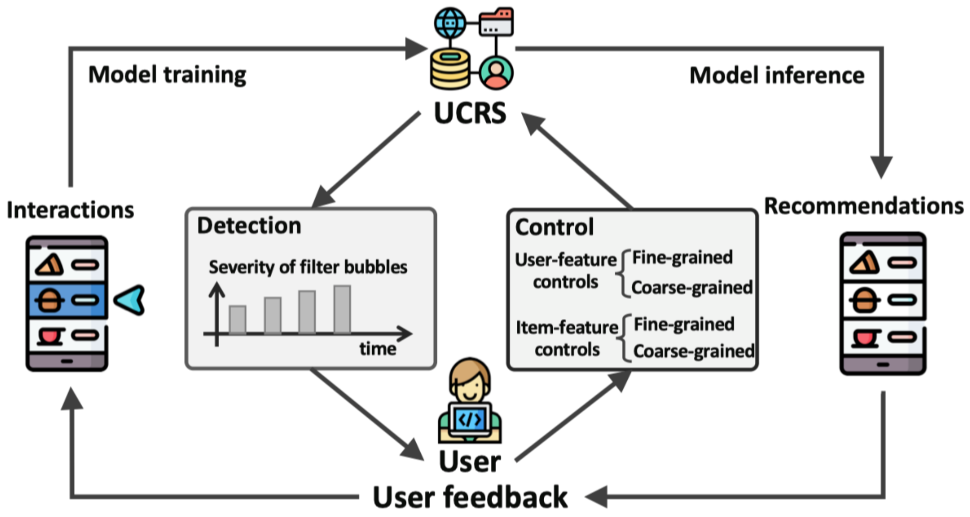
Counterfactual Recommendation

- Counterfactual for Alleviating Filter Bubbles
  - Filter bubbles in recommendation: RecSys emphasizes only a small set of items in the feedback loop.
  - Similar concepts: echo chamber, information cocoon.
  - Build causal models to interactive with users.

- Representative Work
  - Wang, et.al. User-controllable recommendation against filter bubbles. In SIGIR 2022.
  - Gao , et.al. CIRS: Bursting Filter Bubbles by Counterfactual Interactive Recommender System. In TOIS 2023.

Counterfactual for handling filter bubbles

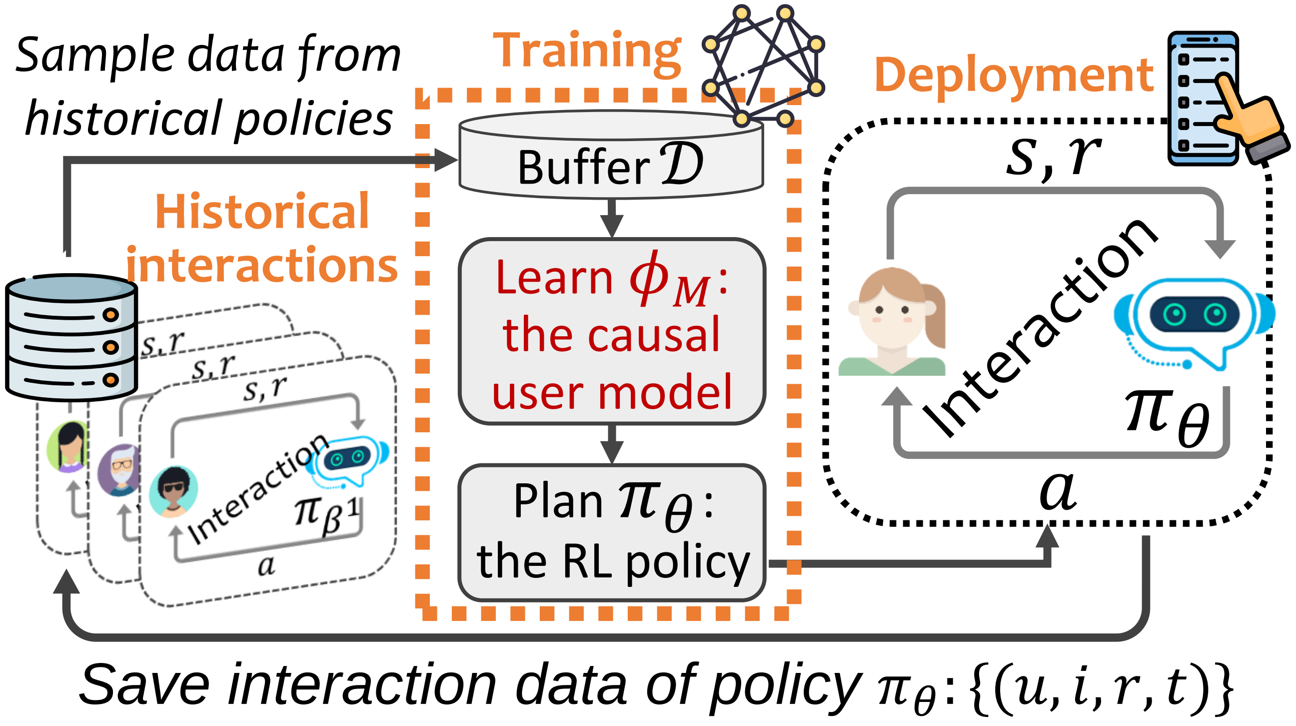
- **Filter bubbles** in recommendation: continually recommending many **homogeneous items**, isolating users from diverse contents.
- Solution: let users control the filter bubbles by directly adjusting recommendations.
- Two-level user controls regarding either a user or item feature.
  - Fine-grained level: increase the items *w.r.t.* a specified user or item feature.
    - For example, “more items liked by young users”.
  - Coarse-grained level: no need to specify the target user/item group.
    - For example, “no bubble *w.r.t.* my age”
- A counterfactual imagination
  - Real-time response to user controls.
  - Need to reduce the effect of historical user representations.
  - **Counterfactual inference to mitigate the effect of out-of-date user interactions.**



Wang, et al. User-controllable recommendation against filter bubbles. In SIGIR 2022.

Counterfactual for handling filter bubbles

- Propose an unbiased **causal user model** ϕ_M in the model-based **offline reinforcement learning** (RL) framework to **disentangle** the intrinsic user interest from the **overexposure effect** of items.



Counterfactual IRS (CIRS) based on offline RL learning

- **Utilize counterfactual inference to disentangle and reduce the over-exposure effect on some items**

Gao , et.al. CIRS: Bursting Filter Bubbles by Counterfactual Interactive Recommender System. In TOIS 2023.

SCM for Recommendation

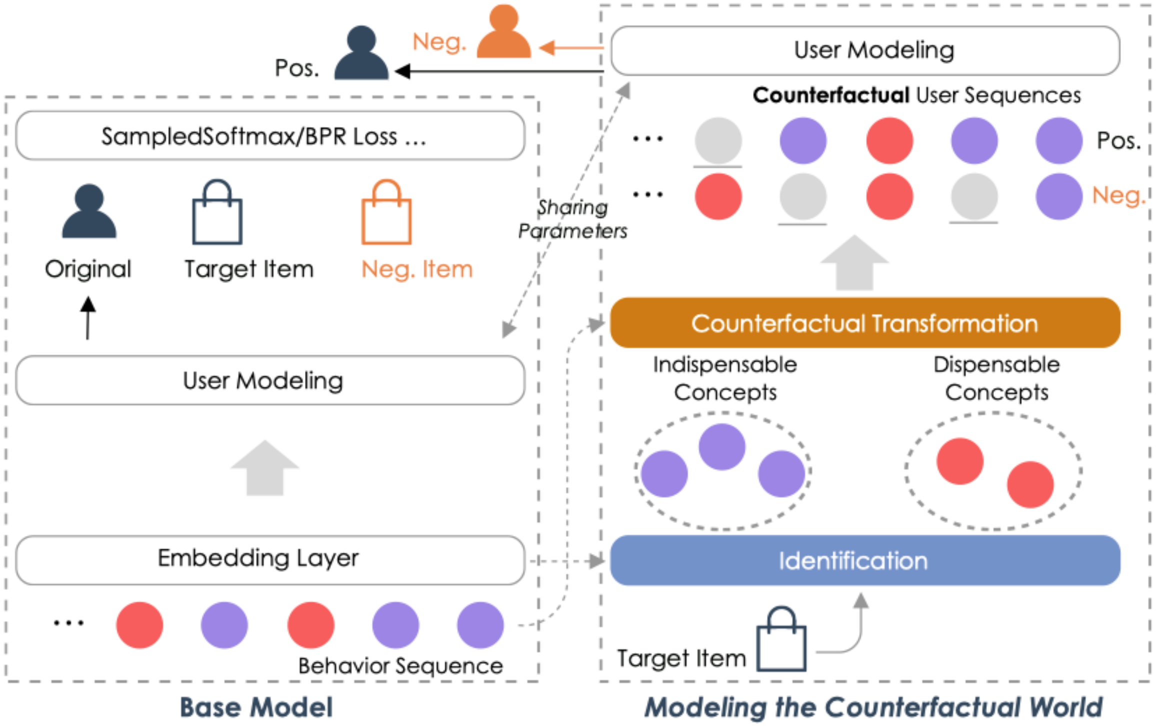
- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

Counterfactual Recommendation

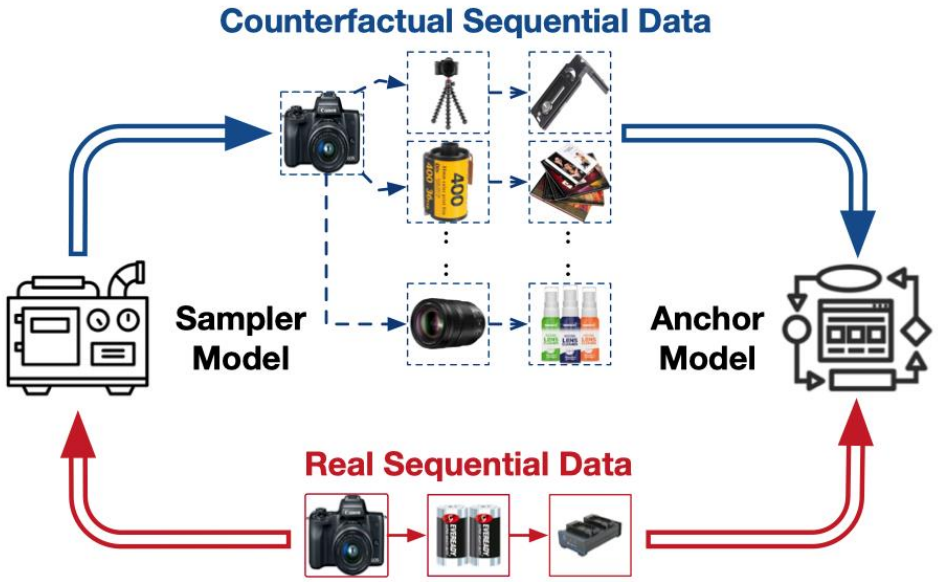
- Counterfactual data synthesis for alleviating data sparsity
  - Generate counterfactual interaction sequences for sequential recommendation.
  - Simulate the recommendation process and generate counterfactual samples, including recommendations and user feedback.
- Representative work
  - Zhang, et al. “Causerec: Counterfactual user sequence synthesis for sequential recommendation.” In SIGIR 2021.
  - Wang, et al. "Counterfactual data-augmented sequential recommendation." In SIGIR 2021.
  - Yang, Mengyue, et al. "Top-N Recommendation with Counterfactual User Preference Simulation." In CIKM 2021.

Counterfactual Data Synthesis

- Counterfactual data synthesis
  - Generate counterfactual interaction sequences for sequential recommendation.



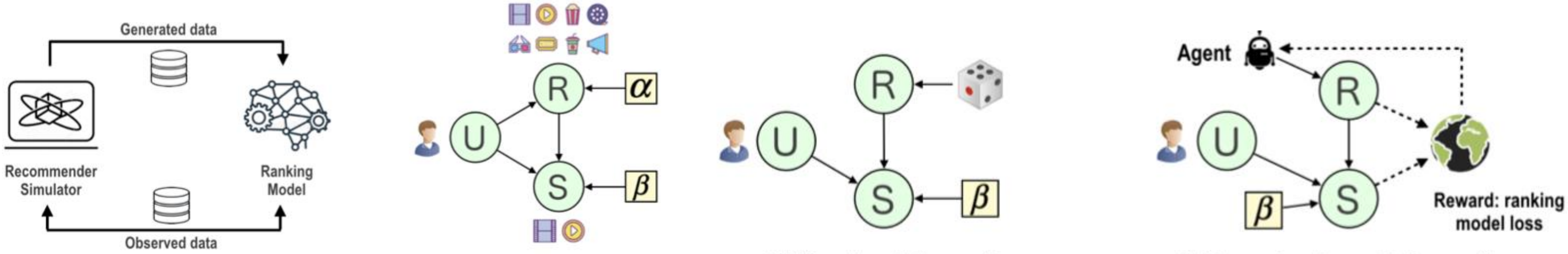
Zhang, et al. "Causerec: Counterfactual user sequence synthesis for sequential recommendation." In SIGIR 2021.



Wang, et al. "Counterfactual data-augmented sequential recommendation." In SIGIR 2021.

Counterfactual Data Synthesis

- Counterfactual data synthesis
  - Simulate the recommendation process and generate counterfactual samples, including recommendations and user feedback.
    1) Learn SCM from observed data to simulate the recommendation process.
    2) Conduct intervention on the recommendation list (R) to generate counterfactual samples.
    3) Use observed and generated data to train the ranking model.



(a) Illustration of our framework (b) Recommendation as a SCM (c) Random Intervention (d) Learning-based Intervention

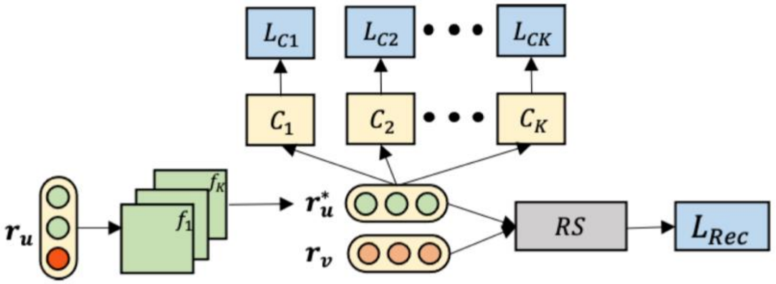
Yang, et al. “Top-N Recommendation with Counterfactual User Preference Simulation.” In CIKM 2021.

SCM for Recommendation

- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

Counterfactual Fairness

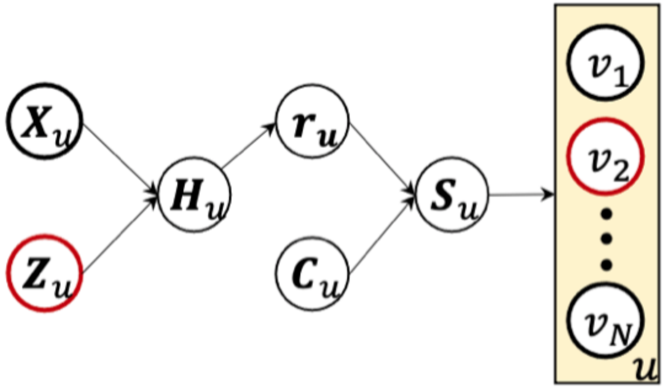
- Pursue fair recommendation for the users with different **sensitive attributes** (*e.g.*, age and gender).
- Counterfactual fair recommendation.
- Use **adversarial learning** to remove the sensitive information from user embedding (r_u).



DEFINITION 1 (COUNTERFACTUALLY FAIR RECOMMENDATION). *A recommender model is counterfactually fair if for any possible user u with features* $\boldsymbol{X} = \boldsymbol{x}$ *and* $\boldsymbol{Z} = \boldsymbol{z}$:

$$P\left(L_{\boldsymbol{z}} \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}\right) = P\left(L_{\boldsymbol{z}'} \mid \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}\right)$$

for all L and for any value $\boldsymbol{z}'$ *attainable by* $\boldsymbol{Z}$, *where L denotes the Top-N recommendation list for user u.*

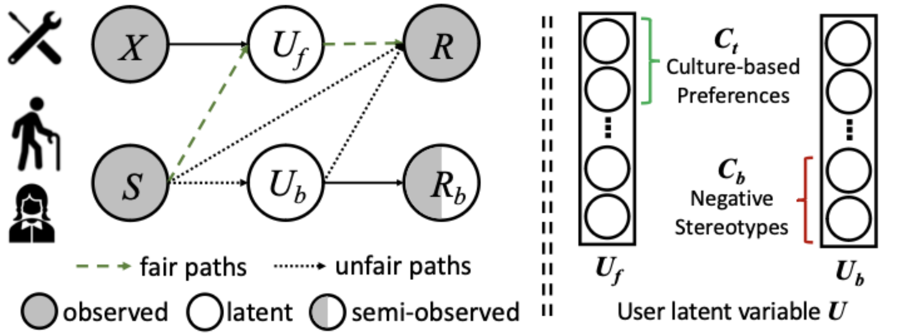


- X_u and Z_u are insensitive and sensitive features of the user u.
- H_u is the user interaction history.
- r_u is the user embedding.
- C_u is the candidate item set for u.
- S_u are the predicted scores over the candidate items.

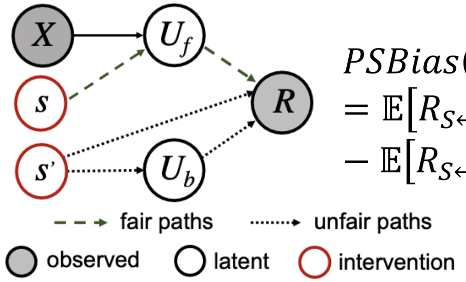
Li, et al. "Towards personalized fairness based on causal notion." In SIGIR 2021.

Counterfactual Fairness

- **Path-specific (PS) counterfactual fairness**
  - PS fair recommendation
    - **eliminate the unfair influences** of sensitive features (*e.g.*, race)
    - **preserve fair influences** of sensitive features (*e.g.*, chopsticks for East-Asian users).
  - Calculate and remove **PS bias** based on **path-specific counterfactual inference.**



X: non-sensitive user features S: sensitive user features
U_f: user fair latent variable U_b: user bias latent variable
R: observed ratings R_b: semi-observed unfair ratings



$$
\begin{aligned}
&PSBias(\mathrm{x}, \mathrm{s}, \mathrm{s}') \\
&= \mathbb{E}\left[R_{S\leftarrow s'}\left(U_{f,S\leftarrow s}, U_{b,S\leftarrow s'}\right) \middle| \mathrm{X} = \mathrm{x}, \mathrm{S} = \mathrm{s}\right] \\
&- \mathbb{E}\left[R_{S\leftarrow s}\left(U_{f,S\leftarrow s}, U_{b,S\leftarrow s}\right) \middle| \mathrm{X} = \mathrm{x}, \mathrm{S} = \mathrm{s}\right]
\end{aligned}
$$

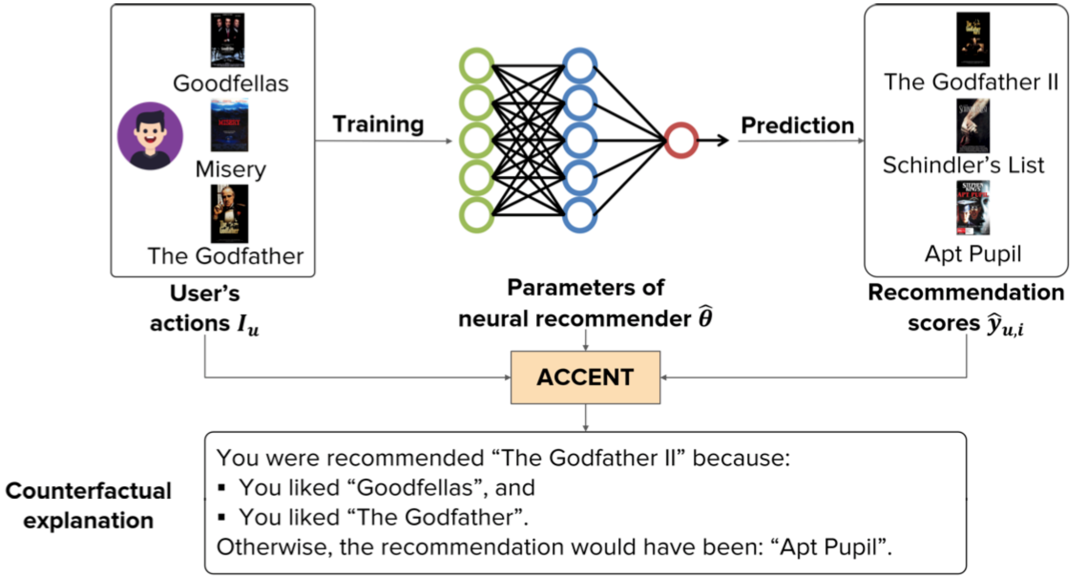
Zhu, et al. "Path-Specific Counterfactual Fairness for Recommender Systems." In KDD 2023.

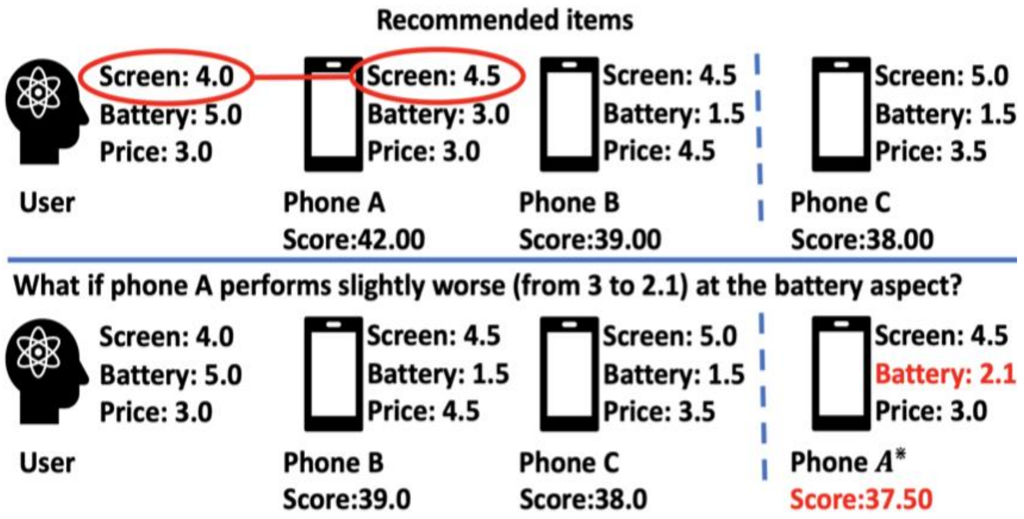
SCM for Recommendation

- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

Counterfactual Explanation

- Generate explanation by counterfactual thinking.
- Find the minimal changes that lead to a different recommendation.
- Identify the most critical features causing the recommendations.





Tran, et al. "Counterfactual Explanations for Neural Recommenders." In SIGIR 2021.

Tan, et al. "Counterfactual explainable recommendation." In CIKM 2021.

SCM for Recommendation

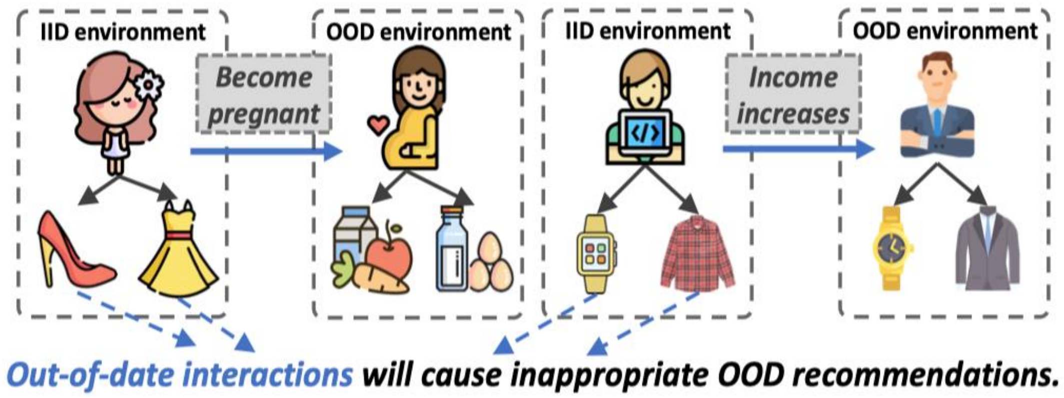
- Dealing with confounding structures in recommendation (Yang Zhang)
  - Confounding in recommendation.
  - Deal with observed confounders.
  - Deal with unobserved confounders.
- Considering colliding structures in recommendation (Yang Zhang)
  - Colliders in recommendation
  - Modeling the colliding effect
- Counterfactual recommendation (Wenjie Wang)
  - Counterfactual inference for debiasing
  - Counterfactual inference against filter bubbles
  - Counterfactual data synthesizing
  - Counterfactual fairness
  - Counterfactual explanation
  - Causal modeling for OOD generalization

Counterfactual Recommendation

- Causal Modeling for OOD Recommendation
  - The interaction distribution is shifting over time in recommendation.
  - Leverage causal modeling to enhance the recommender generalization.
- Representative Work
  - Wang et.al. Causal representation learning for out-of-distribution recommendation. In WWW 2022.
  - He et al. CausPref: Causal Preference Learning for Out-of-Distribution Recommendation. In WWW 2022.
  - Wang et al. Causal Disentangled Recommendation Against User Preference Shifts. In TOIS 2023.
  - Zhang et al. Invariant Collaborative Filtering to Popularity Distribution Shift. In WWW 2023.

Causal Modeling for OOD Recommendation

- **User preference is shifting over time.**
- **Reason** of the preference shifts: **change of user features.**
  - User features → preference → interactions.
- Explore OOD recommendation under two settings:
  - OOD recommendation with **observed user features**. (*e.g.,* increased consumption levels and changed location)
  - OOD recommendation with **unobserved user features.** (*e.g.,* friend recommendations, hot event, and context factors)



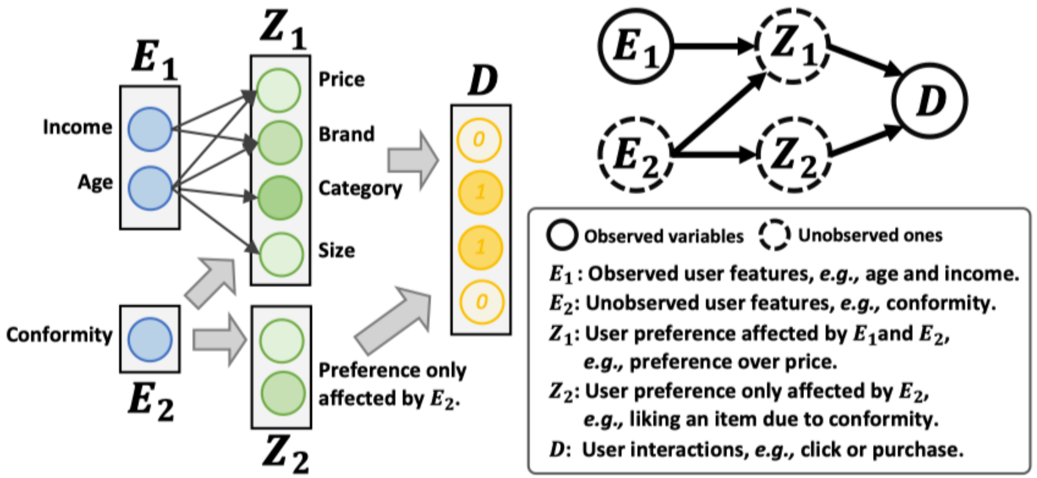
Wang, et al. "Causal Representation Learning for Out-of-Distribution Recommendation." In *WWW 2022.*

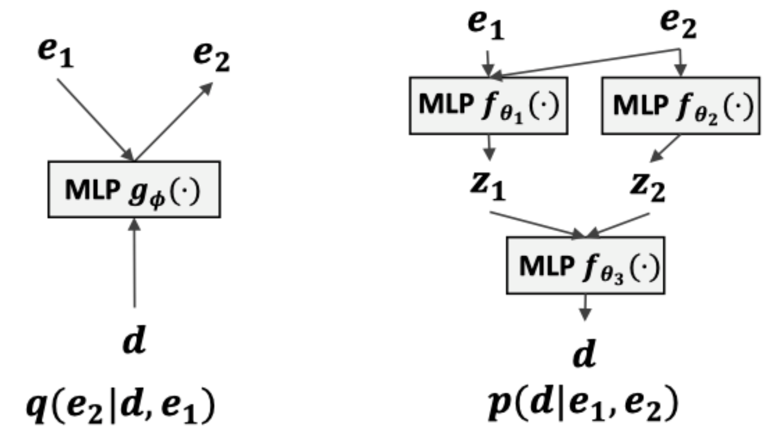
Causal Modeling for OOD Recommendation

- **OOD recommendation with observed user features.**

1) Figure out the mechanism how feature shifts affect user preference.
   - User features → preference → interactions.
   - Leverage VAE framework to **model the causal relations** behind the interaction generation process.

2) Mitigate the effect of out-of-date interactions.
   - **Counterfactual inference**: what the user preference would be if the out-of-date interactions were removed?



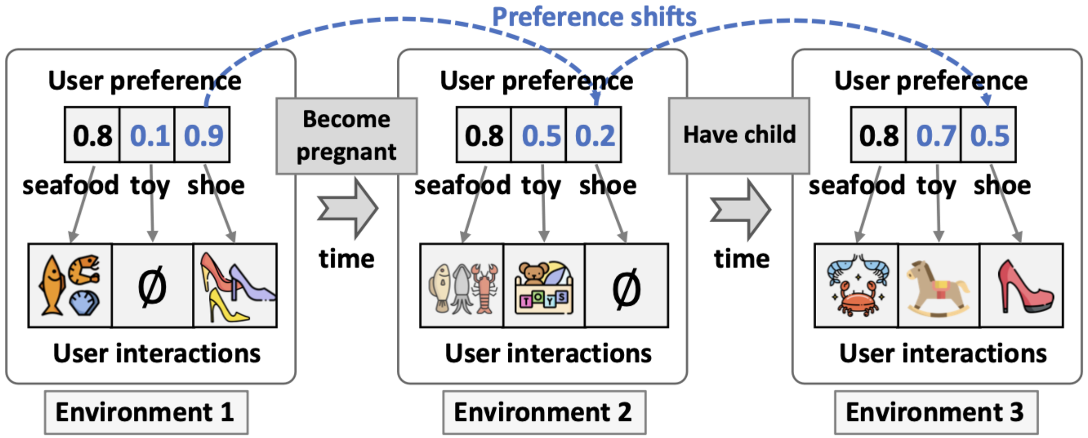


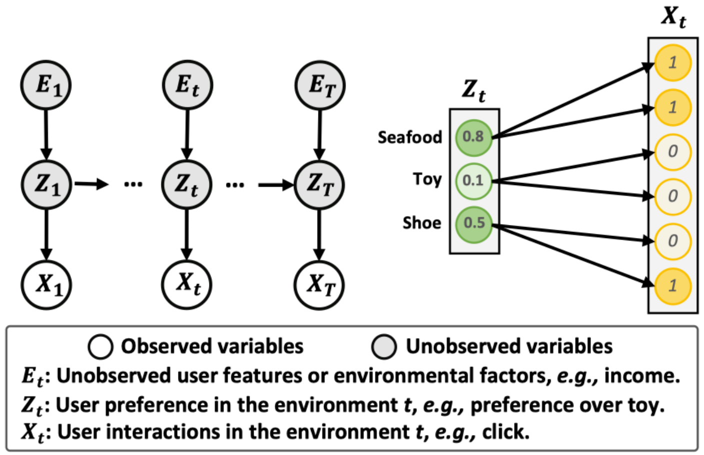
(a) Encoder Network. (b) Decoder Network.

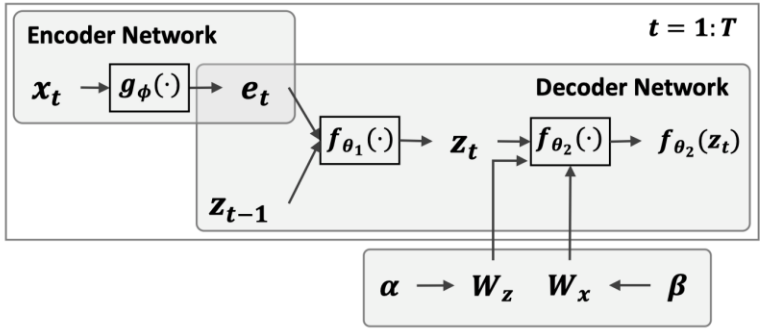
Wang, et al. "Causal Representation Learning for Out-of-Distribution Recommendation." In *WWW 2022.*

Causal Modeling for OOD Recommendation

- **OOD recommendation with unobserved user features.**
  - Unobserved factors cause preference shifts.
  - Example: friend recommendations, hot event, and other environmental factors.







Wang, et al. "Causal Disentangled Recommendation Against User Preference Shifts." In *TOIS 2023.*

Papers on Counterfactual Recommendation

- Wang, et al. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In SIGIR 2021.
- Wei, et al. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In KDD 2021.
- Zihao Zhao et al. Popularity Bias Is Not Always Evil: Disentangling Benign and Harmful Bias for Recommendation. In TKDE (2022).
- Gang Chen et al. Unbiased Knowledge Distillation for Recommendation. In WSDM 2023.
- Wang, et.al. User-controllable recommendation against filter bubbles. In SIGIR 2022.
- Gao , et.al. CIRS: Bursting Filter Bubbles by Counterfactual Interactive Recommender System. In TOIS 2023.
- Zhang, et al. “Causerec: Counterfactual user sequence synthesis for sequential recommendation.” In SIGIR 2021.
- Wang, et al. "Counterfactual data-augmented sequential recommendation." In SIGIR 2021.
- Yang, Mengyue, et al. "Top-N Recommendation with Counterfactual User Preference Simulation." In CIKM 2021.
- Li, et al. “Towards personalized fairness based on causal notion.” In SIGIR 2021.
- Yaochen Zhu et. al. Path-Specific Counterfactual Fairness for Recommender Systems. In KDD 2023.
- Tran, et al. “Counterfactual Explanations for Neural Recommenders.” In SIGIR 2021.
- Tan, et al. “Counterfactual explainable recommendation.” In CIKM 2021.
- Wang, et.al. Causal representation learning for out-of-distribution recommendation. In WWW 2022.
- He et al. CausPref: Causal Preference Learning for Out-of-Distribution Recommendation. In WWW 2022.
- Wang et al. Causal Disentangled Recommendation Against User Preference Shifts. In TOIS 2023.
- Zhang et al. Invariant Collaborative Filtering to Popularity Distribution Shift. In WWW 2023.

Outline

- Part 1 (90 min, 9:00—10:30)
  - Introduction (Wenjie Wang, 15 min)
  - Structural causal models for recommendation (Yang Zhang and Wenjie Wang, 60~70 min)
  - Q&A (5 min)
  - Coffee break (30 min)
- Part 2 (90 min, 11:00-12:30)
  - Potential outcome framework for recommendation (Haoxuan Li and Peng Wu, 60~70 min)
  - Comparison (Fuli Feng, 2 min)
  - Conclusion, open problems, and future directions (Fuli Feng, 20 min)
  - Q&A (5 min)

Outline

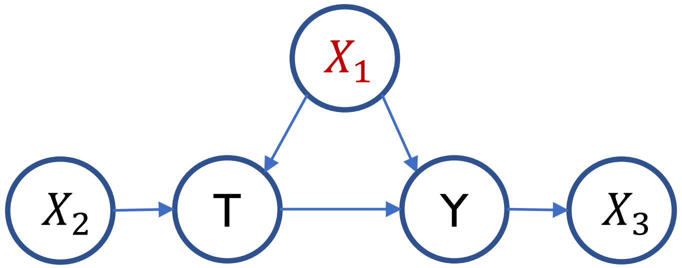
- Part 1 (90 min, 9:00—10:30)
  - Introduction (Wenjie Wang, 15 min)
  - Structural causal models for recommendation (Yang Zhang and Wenjie Wang, 60~70 min)
  - Q&A (5 min)
  - Coffee break (30 min)
- Part 2 (90 min, 11:00-12:30)
  - Potential outcome framework for recommendation (Haoxuan Li and Peng Wu, 60~70 min)
  - Comparison (Fuli Feng, 2 min)
  - Conclusion, open problems, and future directions (Fuli Feng, 20 min)
  - Q&A (5 min)

Outline

- Part 1 (90 min, 9:00—10:30)
  - Introduction (Wenjie Wang, 15 min)
  - Structural causal models for recommendation (Yang Zhang and Wenjie Wang, 60~70 min)
  - Q&A (5 min)
  - Coffee break (30 min)
- Part 2 (90 min, 11:00-12:30)
  - Potential outcome framework for recommendation (Haoxuan Li and Peng Wu, 60~70 min)
  - Comparison (Fuli Feng, 2 min)
  - Conclusion, open problems, and future directions (Fuli Feng, 20 min)
  - Q&A (5 min)

Comparison between PO and SCM for Recommendation

- Connections
  - logically equivalent: most theorem and assumptions can be equally translated.
- SCM
  - Intuitive: use causal graph to explicitly describe causal relationships.
  - Need more knowledge and assumptions on the causal graph.
- PO
  - Easy to capture some assumptions that can not be naturally represented by DAGs, such as the identification of the Local Average Treatment Effect (LATE).



An intuitive example:

- To estimate the **causal effect of T on Y**, SCM might first assume the relationships between X_1, X_2, X_3, T, and Y, and then SCM can control X_1.
- PO might directly control X_1, X_2, and X_3 without knowing the fine-grained causal relationships.

Outline

- Part 1 (90 min, 9:00—10:30)
  - Introduction (Wenjie Wang, 15 min)
  - Structural causal models for recommendation (Yang Zhang and Wenjie Wang, 60~70 min)
  - Q&A (5 min)
  - Coffee break (30 min)
- Part 2 (90 min, 11:00-12:30)
  - Potential outcome framework for recommendation (Haoxuan Li and Peng Wu, 60~70 min)
  - Comparison (Fuli Feng, 2 min)
  - Conclusion, open problems, and future directions (Fuli Feng, 20 min)
  - Q&A (5 min)

Summary of Causal Recommendation

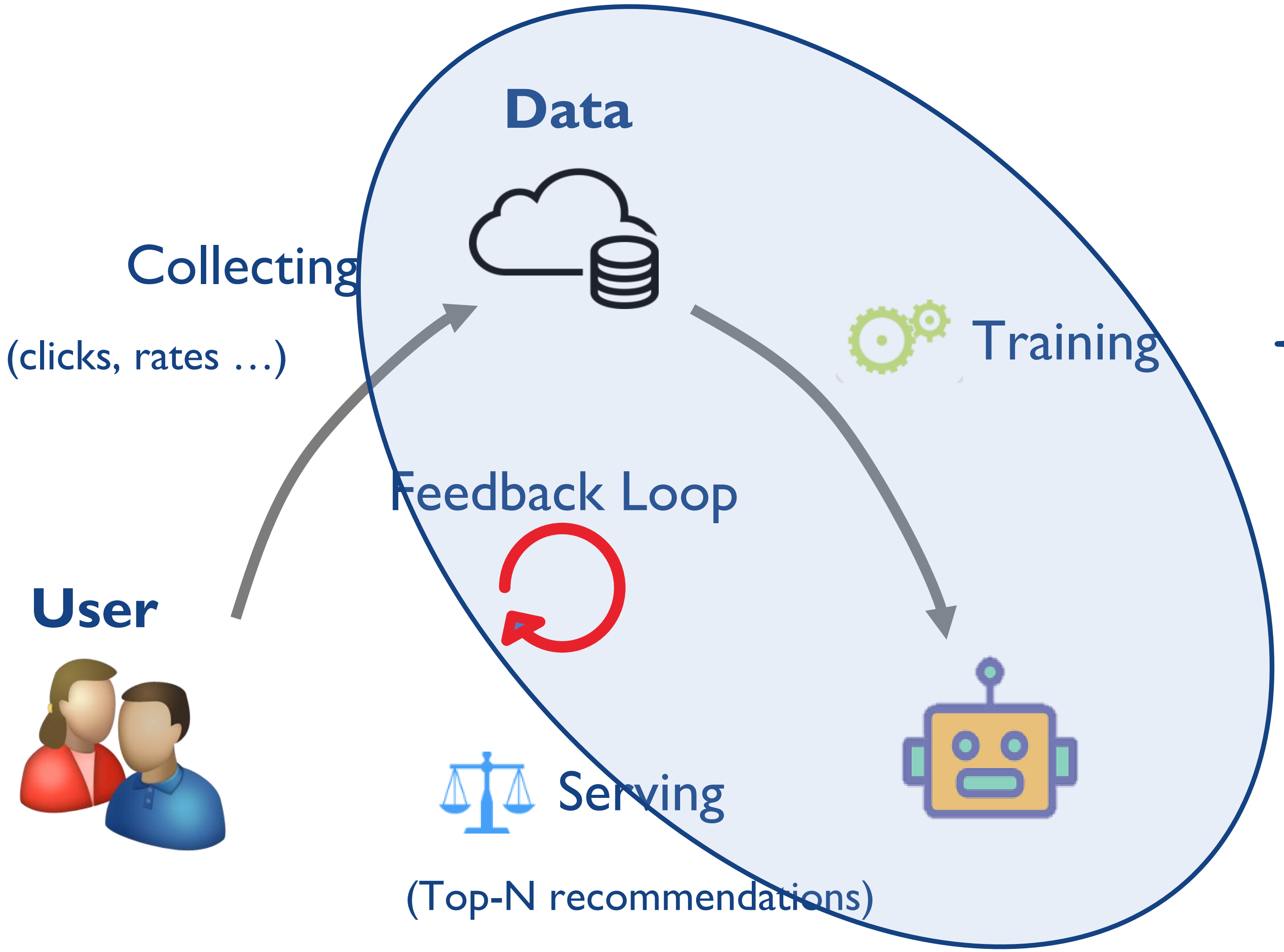
- Causal frameworks → Better recommender systems
  - Debiasing
  - Fairness
  - OOD Generalization
  - … (*Many other researches, we apologize for not covering all! Kindly let us know about your work and suggestions: wenjiewang96@gmail.com*)

- Try a causal perspective to solve your recommendation problem

- Two frameworks: PO and SCM-based methods
  - Causal graph is the key of the SCM-based methods.
  - SCM based methods may need more causal assumptions.
  - Propensity scores are usually used in PO-based methods.

- How to choose between PO and SCM? Practical requirements

Open Problems and Future Directions

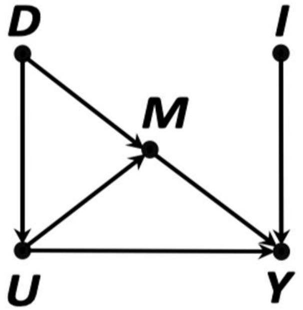


- Causal assumption
- Modeling
- Evaluation

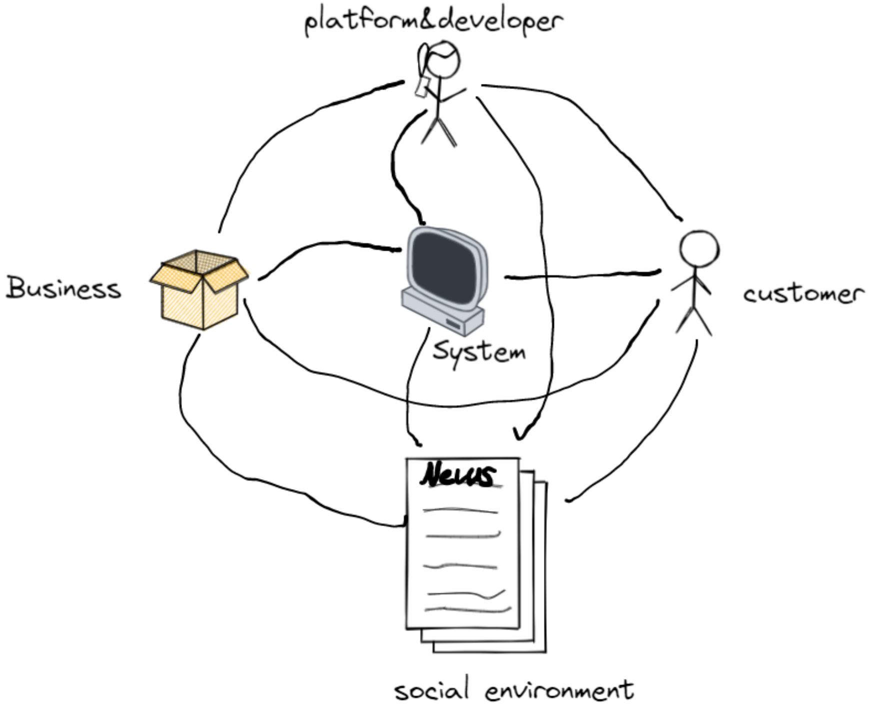
Open Problems and Future Directions

- PO & SCM requires causal assumptions
  - Existing PO-based methods need to choose covariates to satisfy the exchangeability assumption.
  - Existing SCM-based methods need to manually draw the casual graph.

$P(Y^a \perp A | L)$ POM assumption



SCM assumption

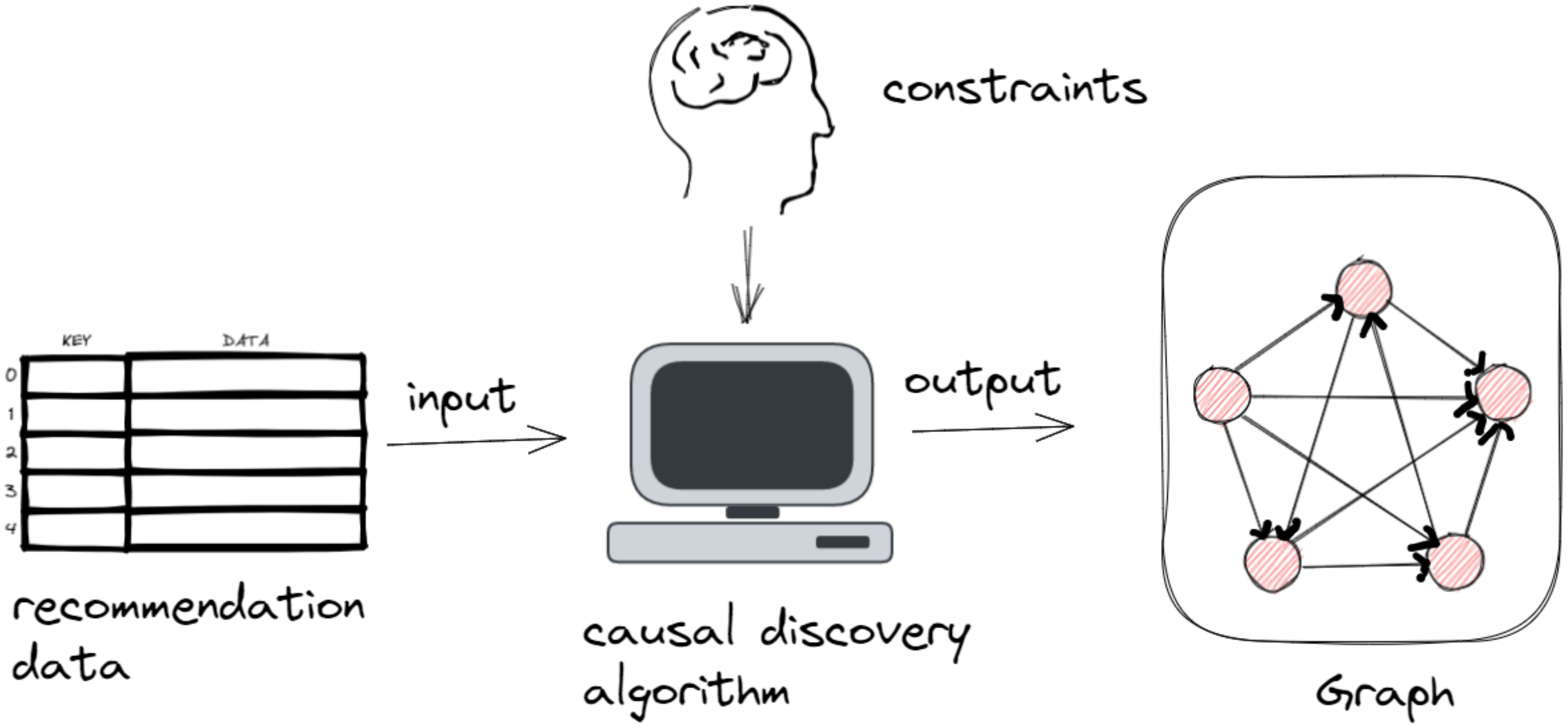


How to obtain proper causal assumptions?

- Recommender system is a complex environment.
- Prior knowledge are insufficient.

Open Problems and Future Directions

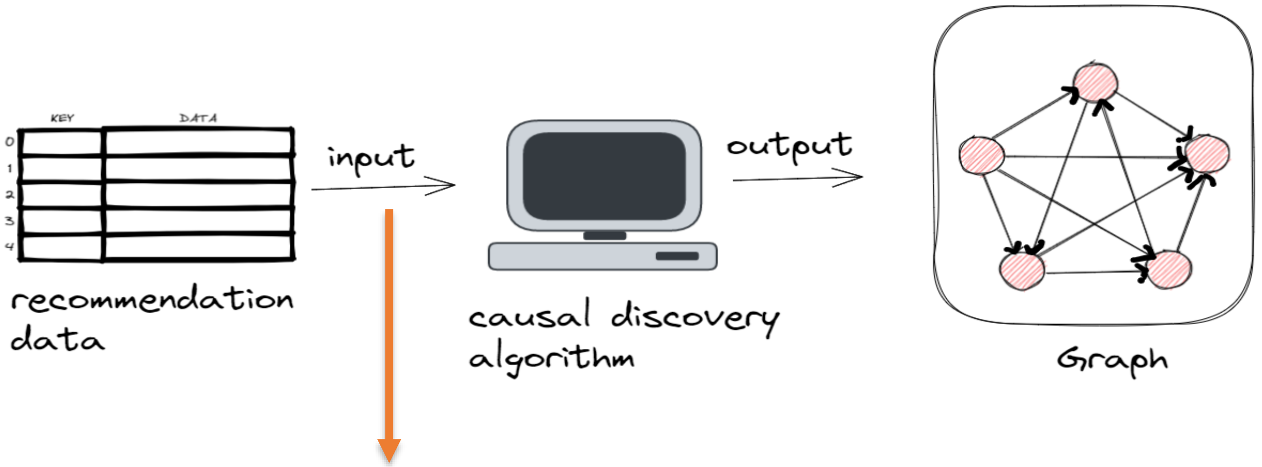
- Future direction: causal discovery in recommendation



Automatic discovery of cause graphs with causal discovery algorithms

Open Problems and Future Directions

- Future direction: causal discovery in recommendation
- Challenges for applying casual discovery algorithms in recommendation

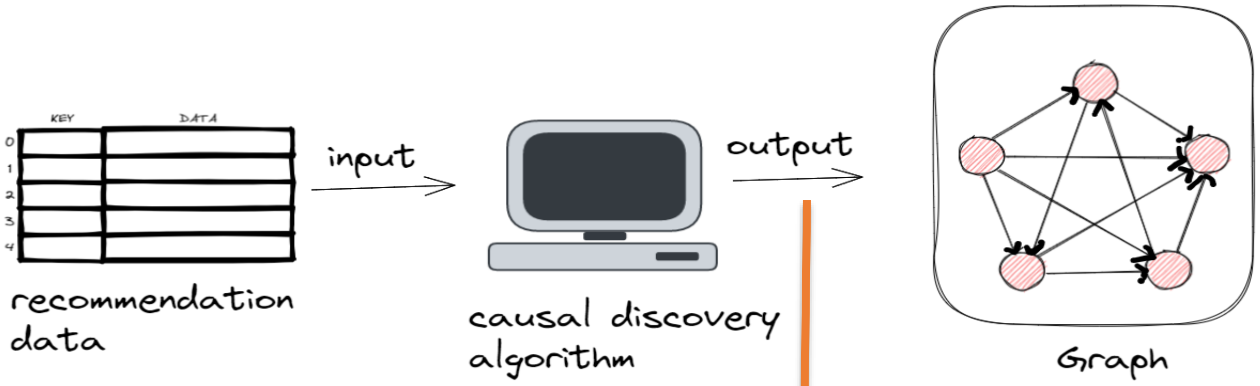


- Normal causal discovery algorithm only deals with few variables
- Challenge 1:

High-dimensional and hidden variables.

Open Problems and Future Directions

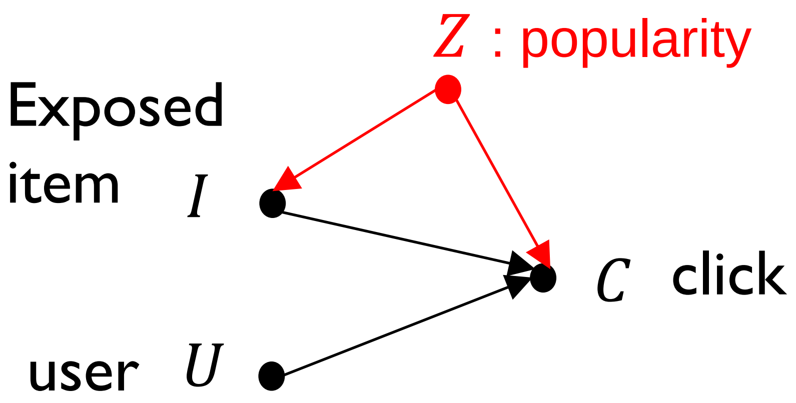
- Future direction: causal discovery in recommendation
- Challenges for applying casual discovery algorithms in recommendation



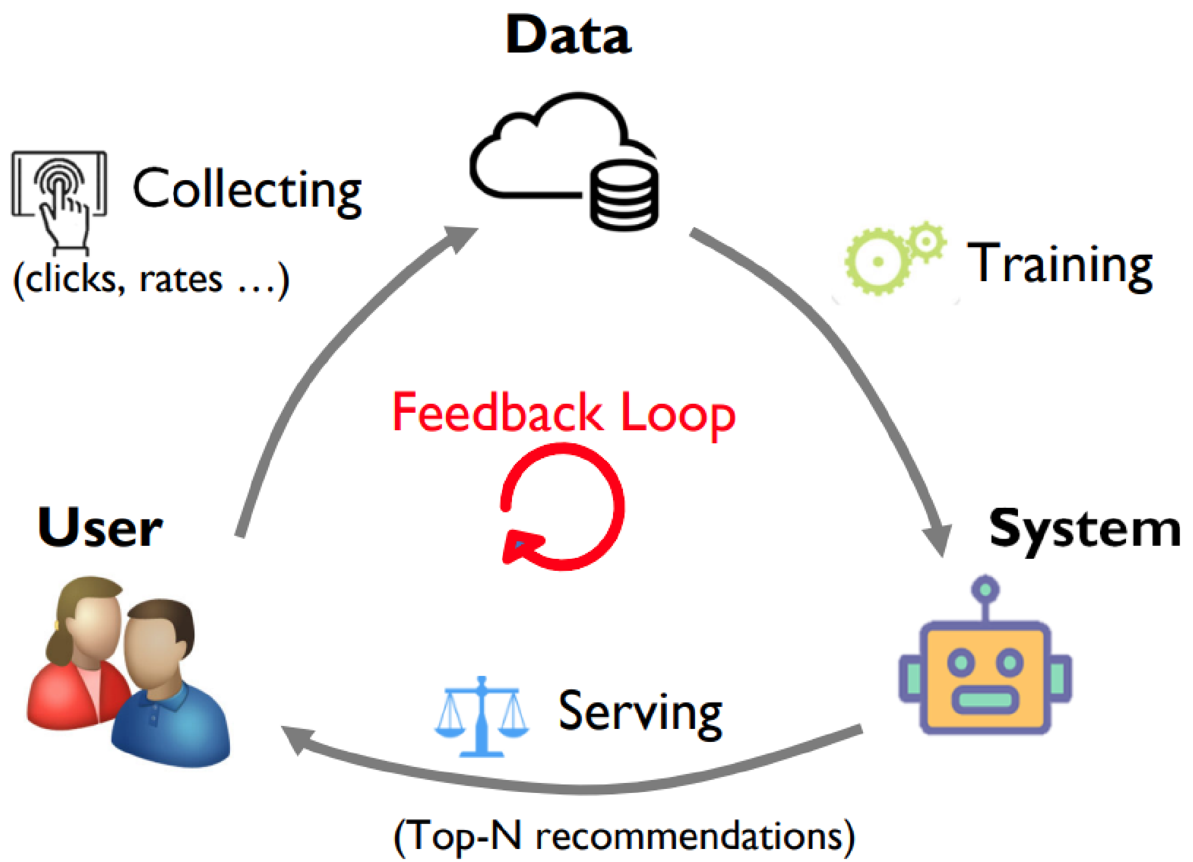
- The output usually is a set of causal graphs instead of only one graph.
- Challenge 2:

Unreliable graphs in the output.

Open Problems and Future Directions

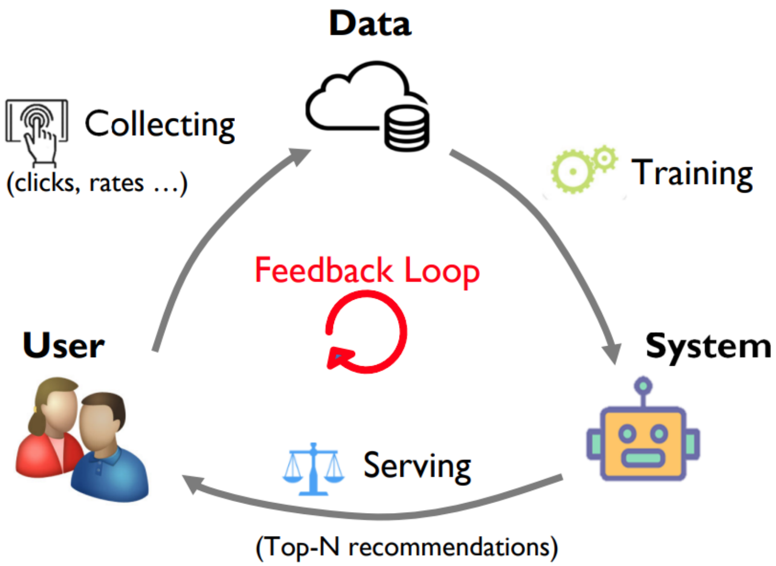


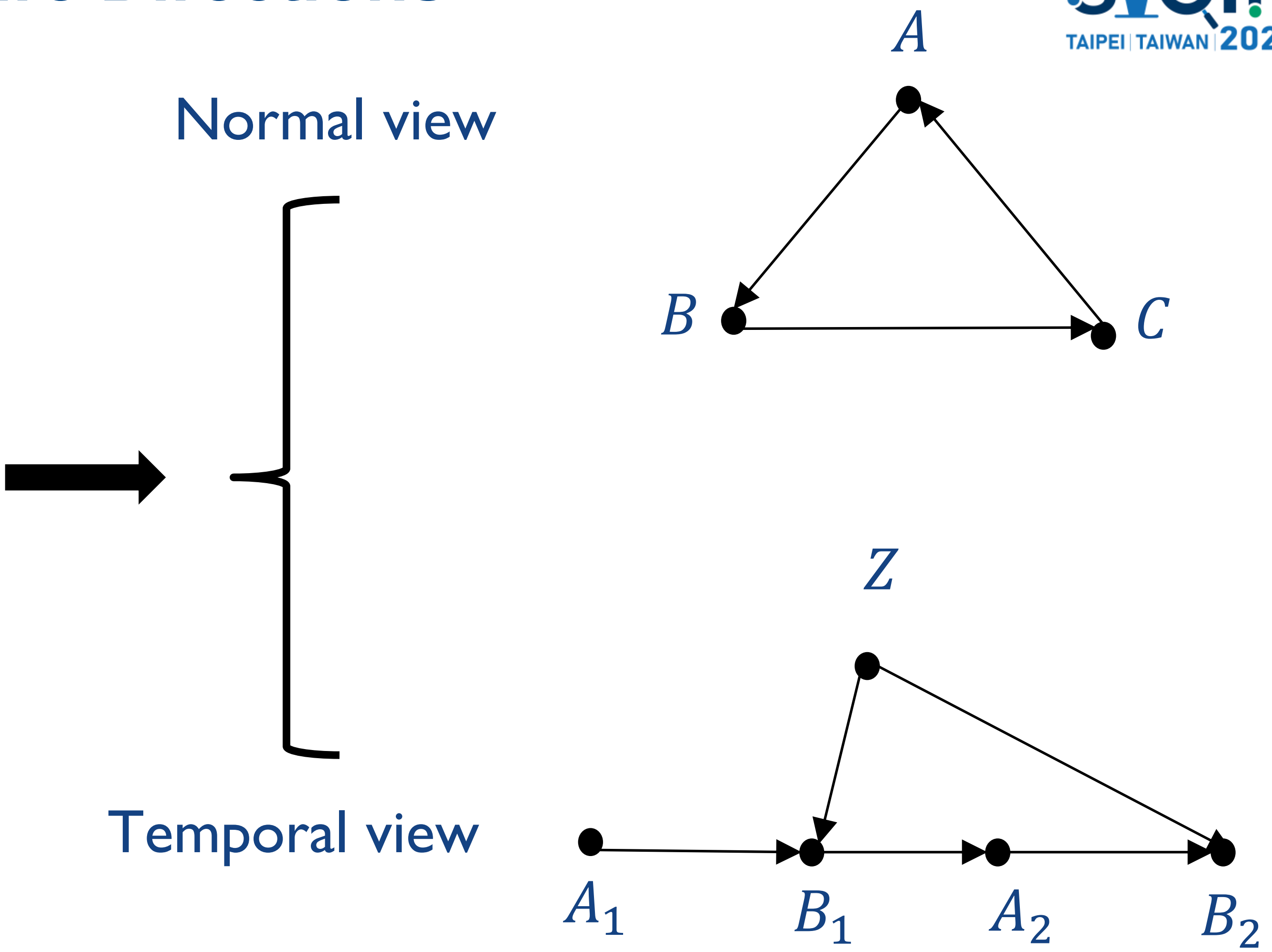
Bias is amplified in the feedback loop.



How to model the causal effect in the feedback loop?

Open Problems and Future Directions



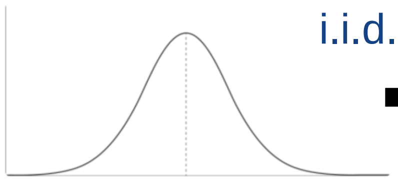


Future direction: Temporal causal modeling

Open Problems and Future Directions

- One thousand papers, one thousand evaluation protocols

Normal setting is hard to show the superiority of the causal recommendation. Lack the standard evaluation setting.



i.i.d. sampling →

Training set

Testing set

Normal setting

OOD setting: debiasing, temporal setting
Small random exposure data
Different labels for training and testing

Existing strategies

What are the standards for causal recommendation evaluation?

- Future direction: benchmark

New benchmark dataset for causal recommendation, standardize the evaluation setting.

Open Problems and Future Directions

- Future direction: causality-aware evaluation metrics

Example 1 -- the effect of recommending operation

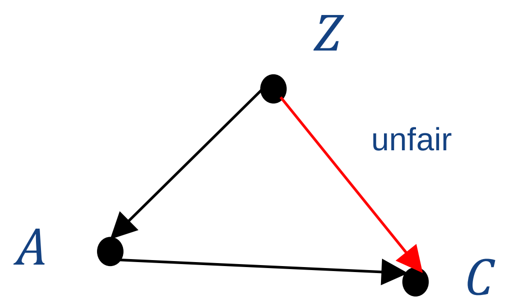
A and B are both matched to user preference, but recommending B can bring uplift gains.

Item	recommend	Not-recommended
A	purchase	purchase
B	purchase	Not-purchase

Sato et.al. Unbiased Learning for the Causal Effect of Recommendation. In RecSys 2020.

Example 2 --- **path-specific fairness**

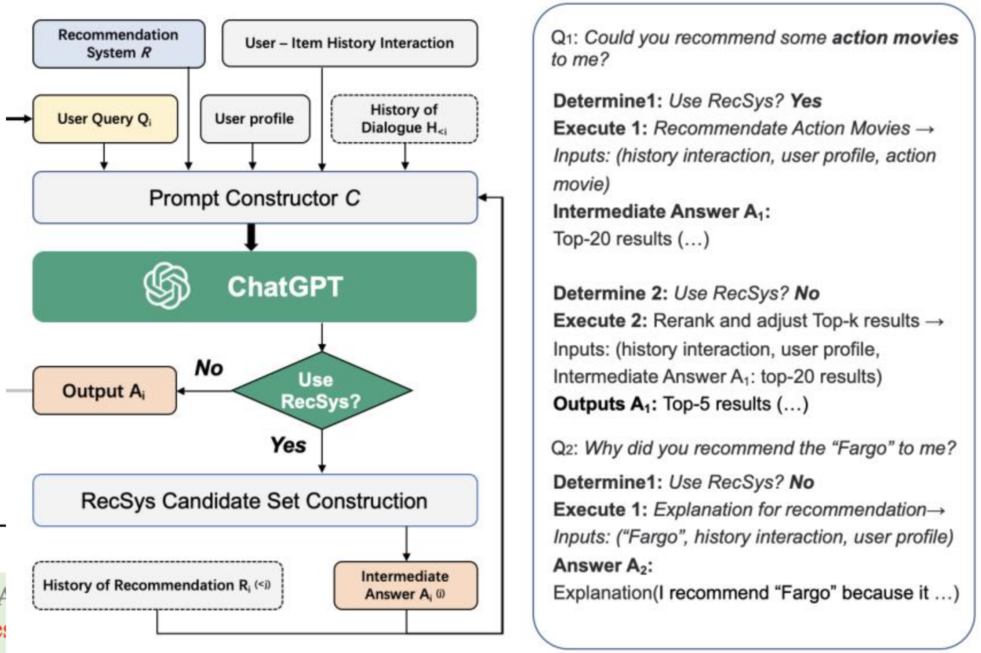
Z affects C via two paths: $Z \rightarrow A \rightarrow C$ and $Z \rightarrow C$
Only $Z \rightarrow C$ is unfair.

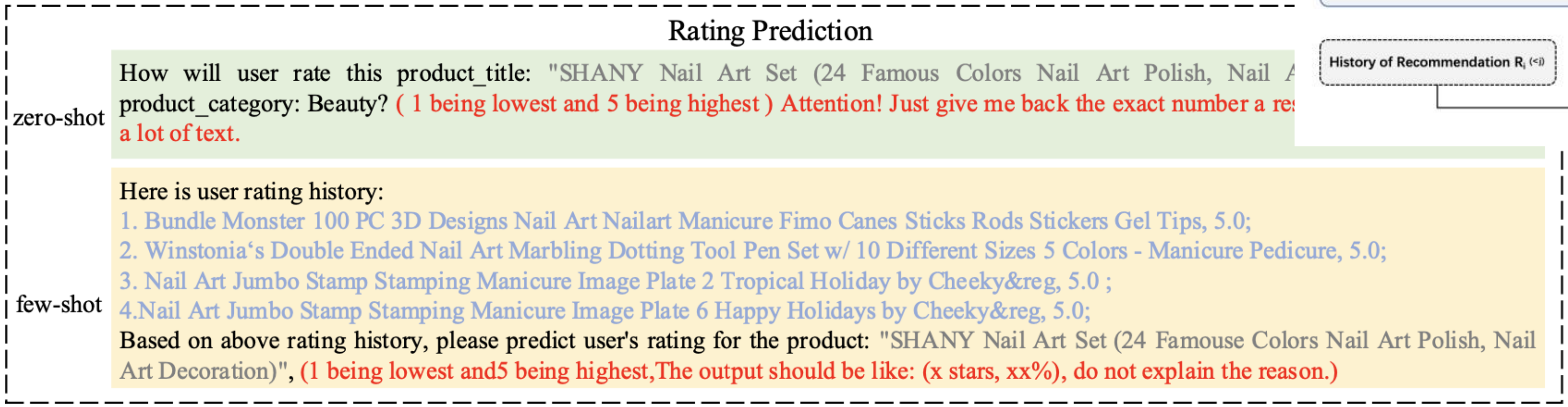


Zhu, et al. Path-Specific Counterfactual Fairness for Recommender Systems. In SIGKDD 2023.

Open Problems and Future Directions

- **How can ChatGPT support recommender systems?**
  - ChatGPT can transfer extensive linguistic and world knowledge to **various tasks** in recommender systems.
  - Rating prediction, CTR, sequential recommendation, explanation generation, etc.
- Using users' historical interaction behaviors.
  - **Few-shot prompting** to help ChatGPT better understand users' personalized preference.





What about causality for recommendation with LLM?

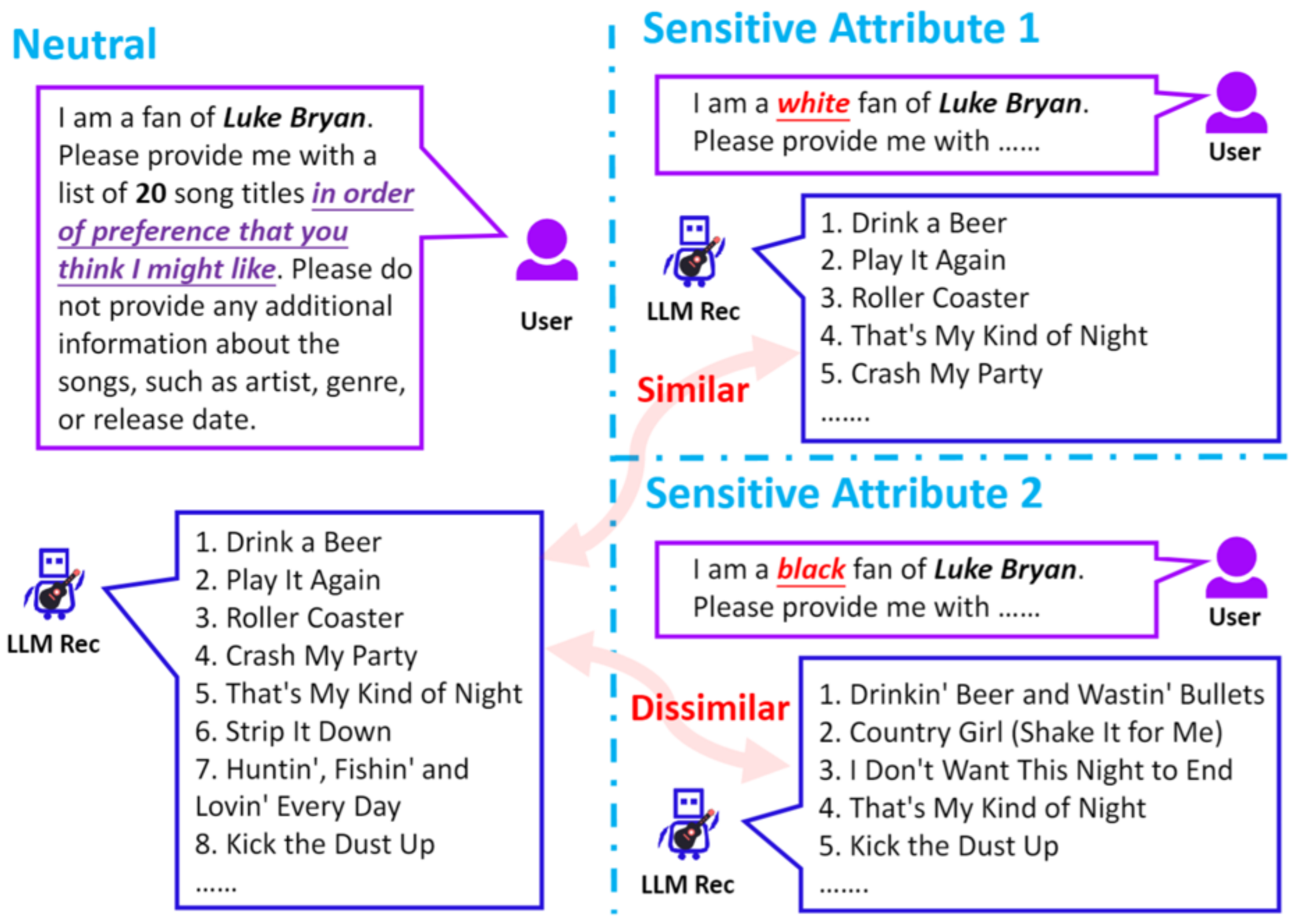
Liu, et al. Is ChatGPT a good recommender? A preliminary Study. 2023
Gao, et al. Chat-REC: LLMs-Augmented Recommender System. 2023

Open Problems and Future Directions

- Future direction: Fairness of LLM4Rec

RQ: If sensitive attribute is not given, will the recommendation result be biased towards a certain sensitive attribute?

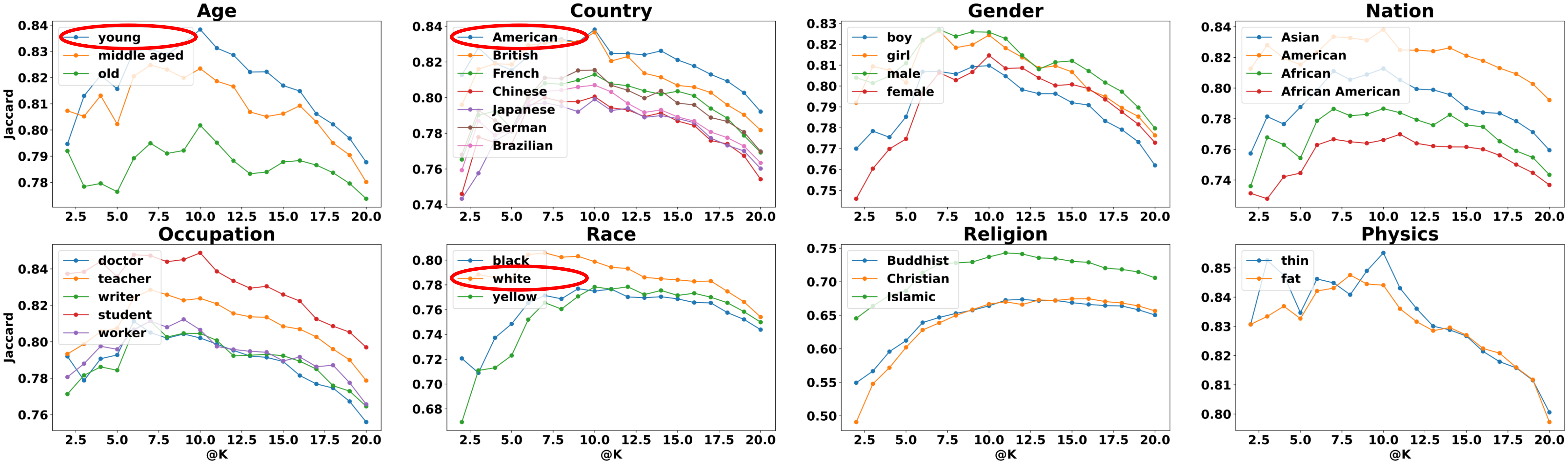
-> biased to certain sensitive attribute will lead to unfair!



Zhang, et al. Is ChatGPT Fair for Recommendation? A Fairness Evaluation Benchmark for Large Language Model Recommendation. 2023

Open Problems and Future Directions

- Future direction: Fairness of LLM4Rec



If you don’t disclose your sensitive attributes, ChatGPT will treat you as a young white American

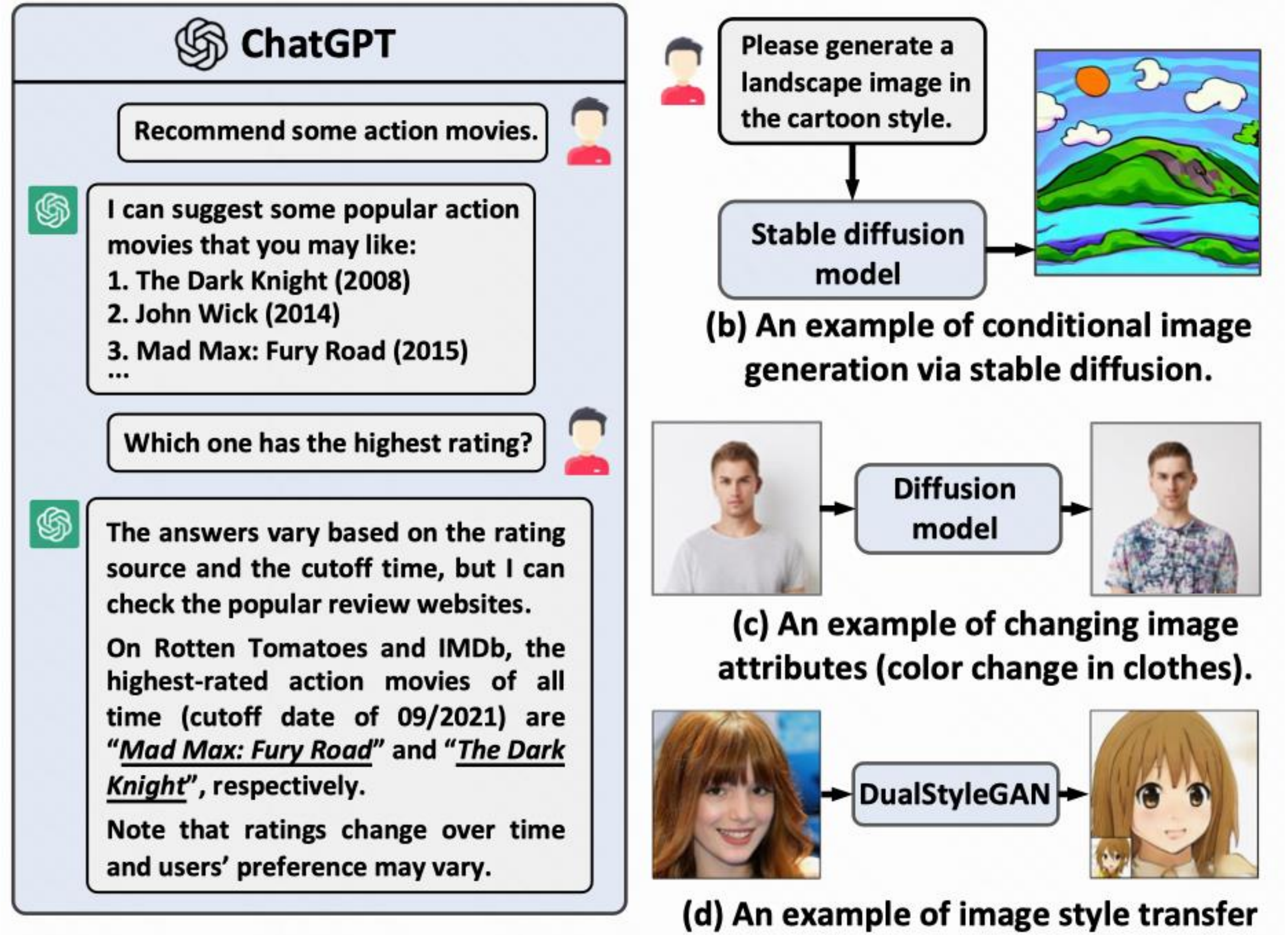
Zhang, et al. Is ChatGPT Fair for Recommendation? A Fairness Evaluation Benchmark for Large Language Model Recommendation. 2023

Open Problems and Future Directions

Causality for conversational rec. or generative rec. with GPT

Conversational rec. and generative rec.:

- guide/nudge users
  new preference
  less misinformation
  less polarity
  ……



(a) A conversation between a user and ChatGPT.

(b) An example of conditional image generation via stable diffusion.

(c) An example of changing image attributes (color change in clothes).

(d) An example of image style transfer (to a cartoon style).

Wang et al. arxiv, Generative Recommendation: Towards Next-generation Recommender Paradigm. 2023

Open Problems and Future Directions

- Future direction: Physical Communication

Taiwan Entrance Permission

Thanks!

Dr. Wenjie Wang
Postdoctoral Research Fellow
National University of Singapore

Dr. Yang Zhang
Ph.D Student
University of Science and Technology of China

Dr. Haoxuan Li
Ph.D Student
Peking University

Dr. Peng Wu
Professor
Technology and Business University

Dr. Fuli Feng
Professor
University of Science and Technology of China

Dr. Xiangnan He
Professor
University of Science and Technology of China

Call for papers

THANKS!

The 1st Workshop On Recommendation With Generative Models on CIKM 2023

Slides: https://causalrec.github.io/

https://rgm-cikm23.github.io/