CSDS 452 Causality and Machine Learning

Lecture 13: Causal effect in applications

Instructor: Jing Ma

Fall 2024, CDS@CWRU

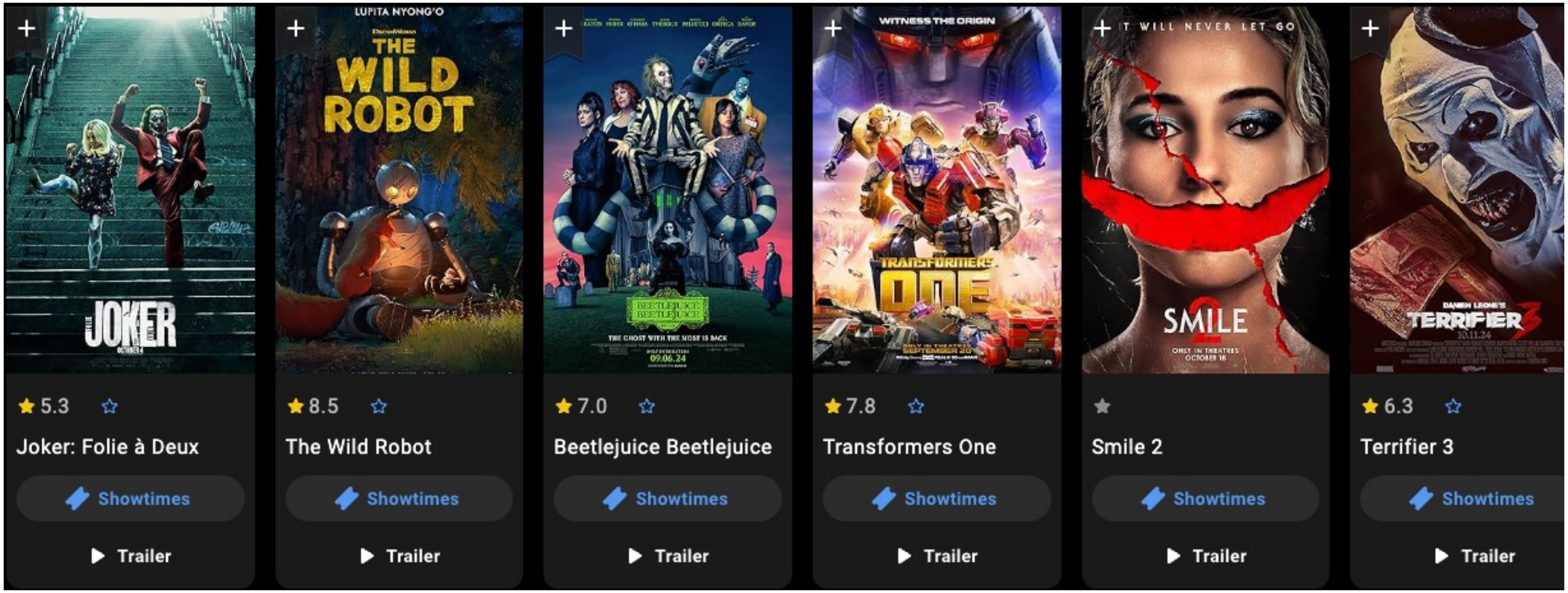
Causality can be applied everywhere

- Recommendation
- Biology and healthcare
- Political science
- Economics
- ……

Outline

- Causal inference in recommender system
  - Recommendation as treatment
  - Biases in recommendation: a causal view
- Other directions

Movie Recommendation



What movies do you like?

How does movie recommendation work?

A/B Testing

- 1. Select the treatments
  - Recommend movie A
  - Recommend movie B
- 2. Randomly assign users
  - Control group: receive recommendation A
  - Treated group: receive recommendation B

The data we want

User-item rating matrix: U x I

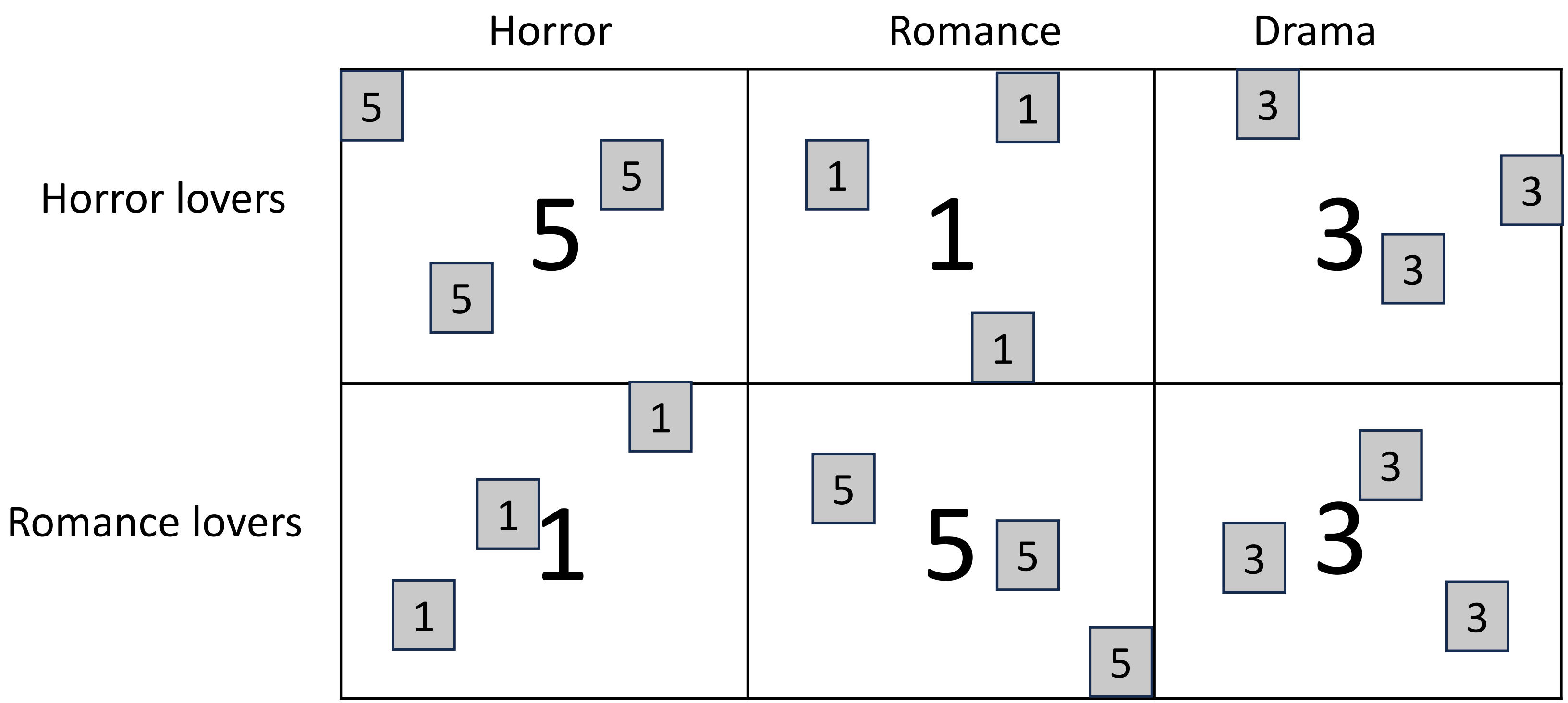
All observed!

	Horror	Romance	Drama
Horror lovers	5	1	3
Romance lovers	1	5	3

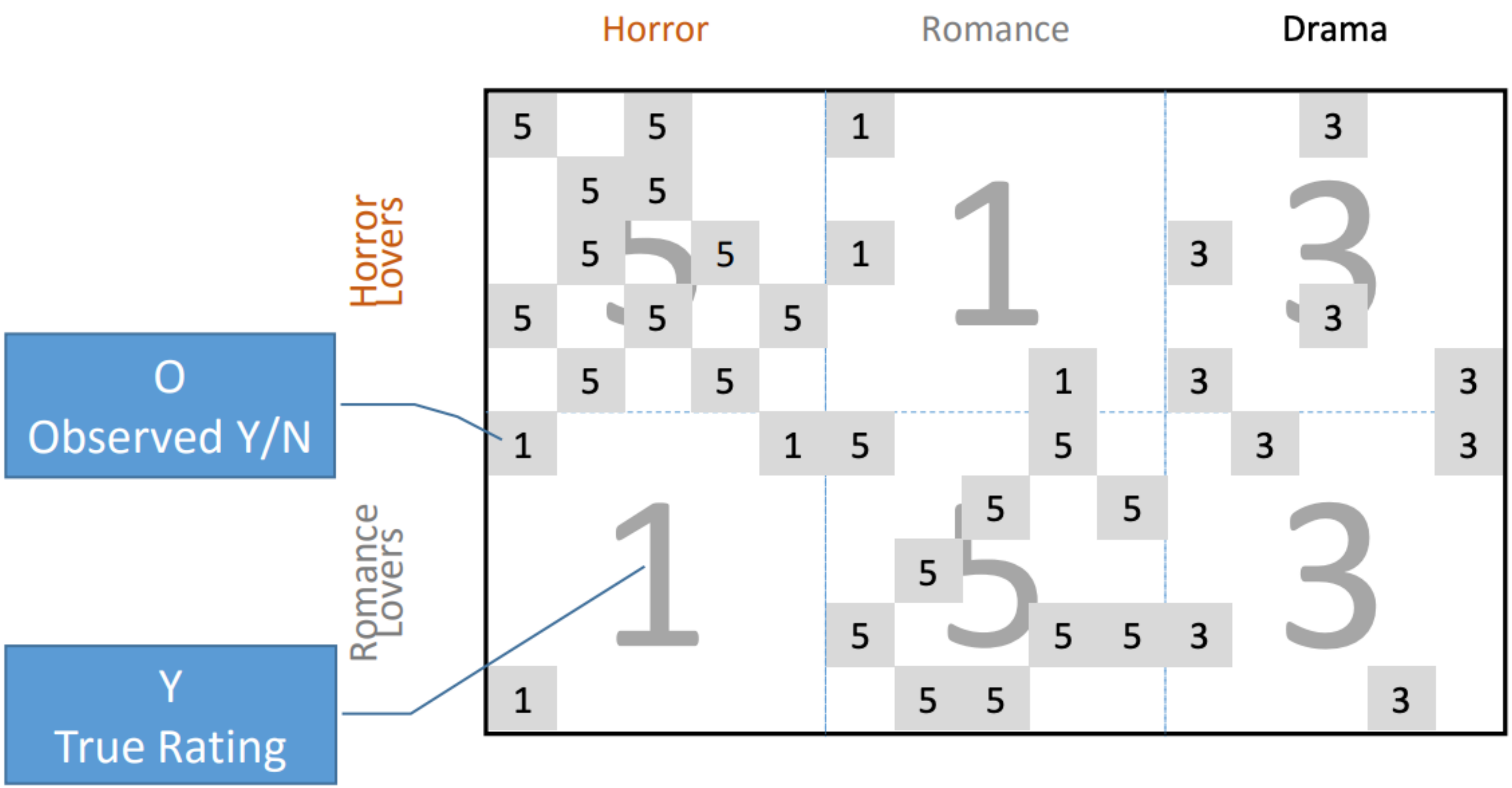
The data that is not bad

User-item rating matrix: U x I

Data missing at random



However, the data we have

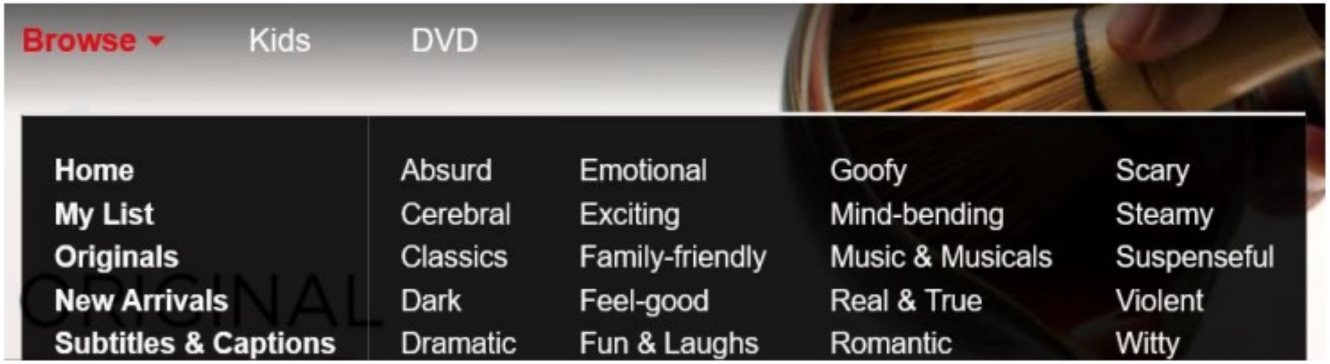


Data is Missing Not At Random (MNAR)

Example adapted from (Steck et al., 2010)

Selection Bias in Recommendation

- Why is there selection bias?
  - User-induced bias (e.g., browsing)



  - System-induced bias (e.g., advertising)

Challenges

- Missing data:
  selection bias or confounding bias.

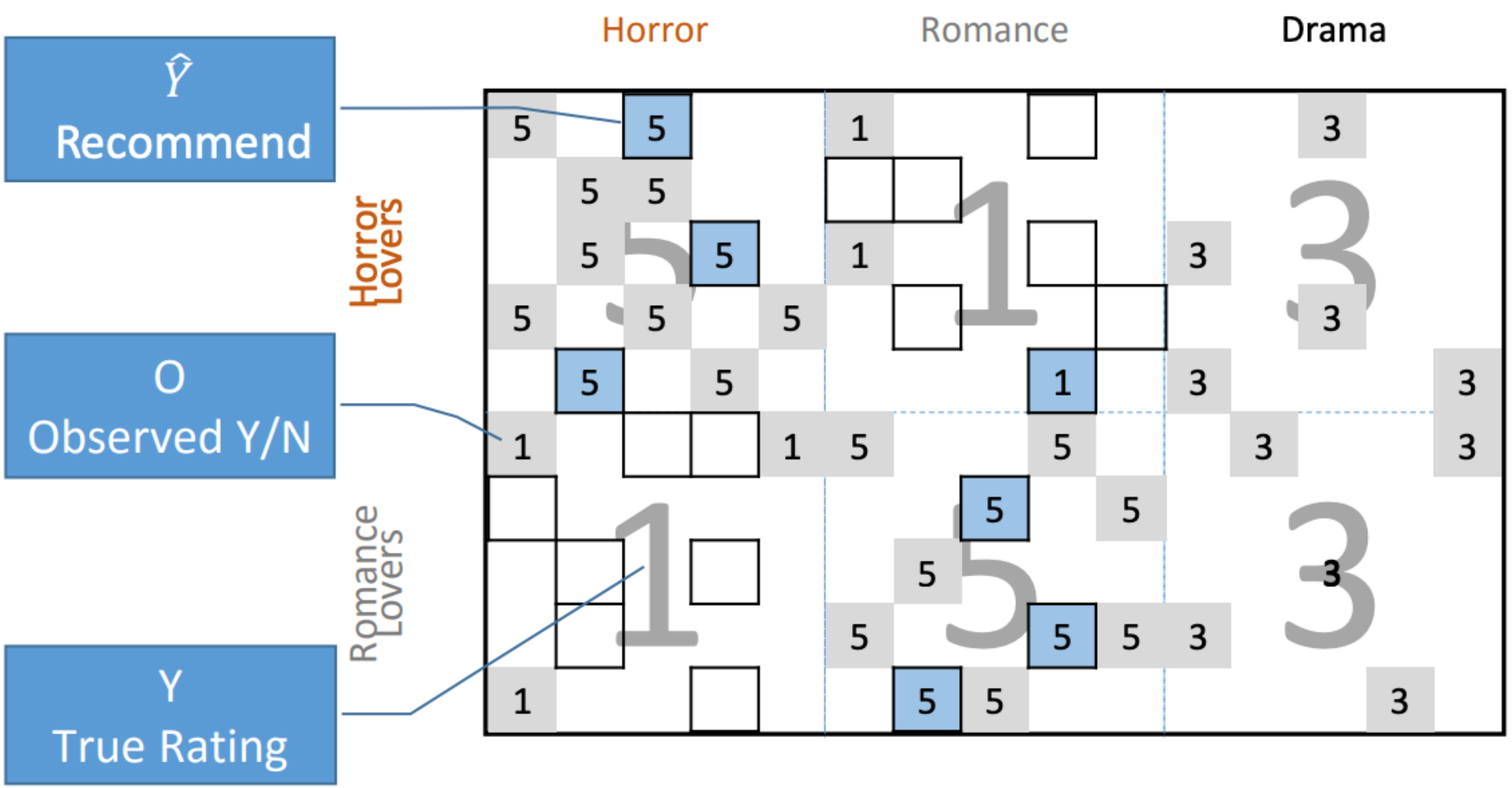
- Data sparsity:

	ML 100K	Coat Shopping	Yahoo! R3
#users	943	290	15400
#items	1682	300	1000
#MNAR ratings	100000	6960	311704
#MAR ratings	0	4640	54000

Missing rate is very high:

- ML 100K: $100000/(943 * 1682) = 0.063$;
- Coat Shopping: $6960/(290 * 300) = 0.080$;
- Yahoo! R3: $311704/(15400 * 1000) = 0.020$.

Evaluating Recommendations under Selection Bias



Observed ratings are misleading due to selection bias

Evaluating recommendations under Selection Bias

$\hat{Y}_1$	$\hat{Y}_2$

Which prediction is better?

$\hat{Y}_1$

	Horror					Romance					Drama				
Horror Lovers	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
Romance Lovers	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3

$\hat{Y}_2$

	Horror					Romance					Drama				
Horror Lovers	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
Romance Lovers	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5

Evaluating Predicted Ratings under Selection Bias

$\hat{Y}_1$
Pred Ratings (worse)

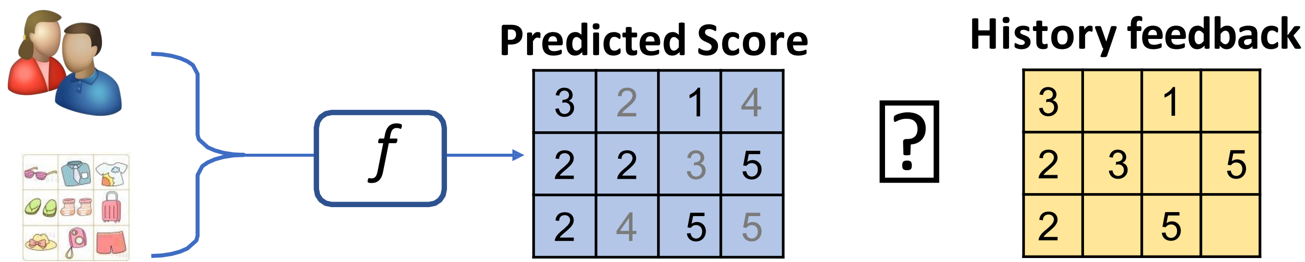
	Horror					Romance					Drama				
Horror Lovers	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
Romance Lovers	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3
	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3

$\hat{Y}_2$
Pred Ratings (better)

	Horror					Romance					Drama				
Horror Lovers	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
	5	5	5	5	5	1	1	1	1	1	5	5	5	5	5
Romance Lovers	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5
	1	1	1	1	1	5	5	5	5	5	5	5	5	5	5

Machine learning joins the game!

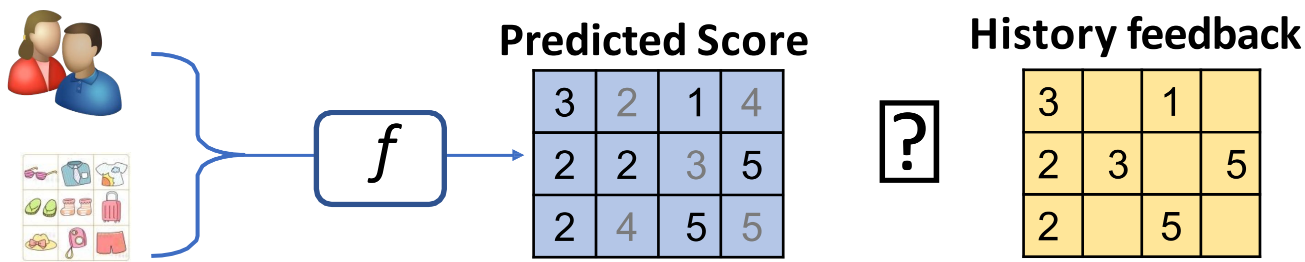
- **Mainstream Models: Fitting Historical Data**
- Minimizing the difference between historical feedback and model prediction



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

Machine learning joins the game!

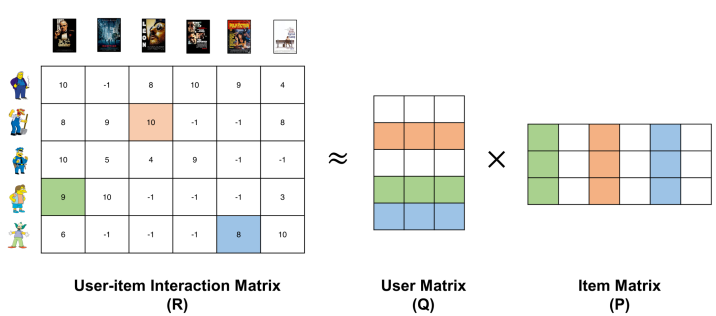
- **Mainstream Models: Fitting Historical Data**
- Minimizing the difference between historical feedback and model prediction



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

Shallow representation learning

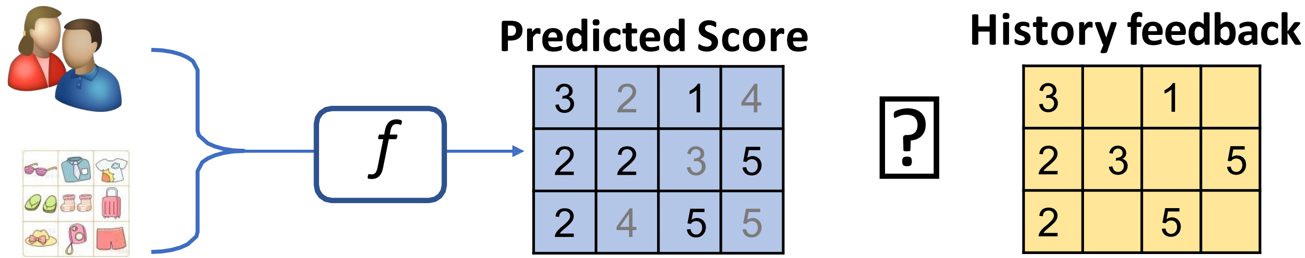
- Matrix factorization & factorization machines



Machine learning joins the game!

- **Mainstream Models: Fitting Historical Data**

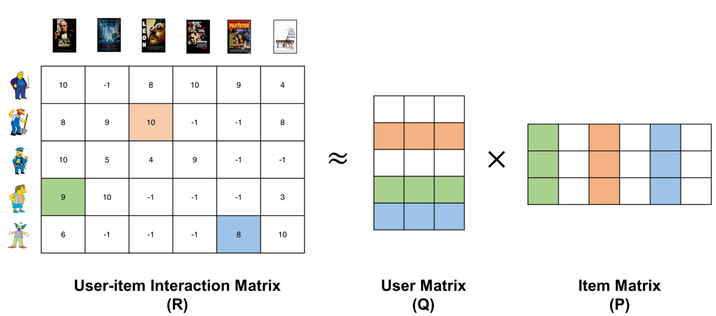
- Minimizing the difference between historical feedback and model prediction



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

Shallow representation learning

- Matrix factorization & factorization machines



Neural representation learning

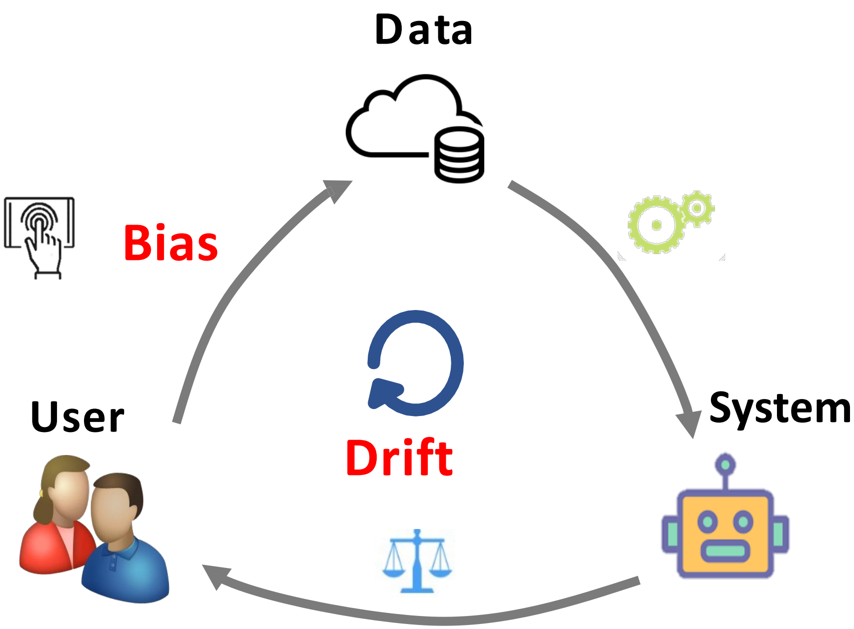
- Neural collaborative filtering
- Graph neural networks
- Sequential model
- Textual & Visual encoders

Learning **correlations** between input features and interaction labels

Shortcomings of Data-Driven Methods

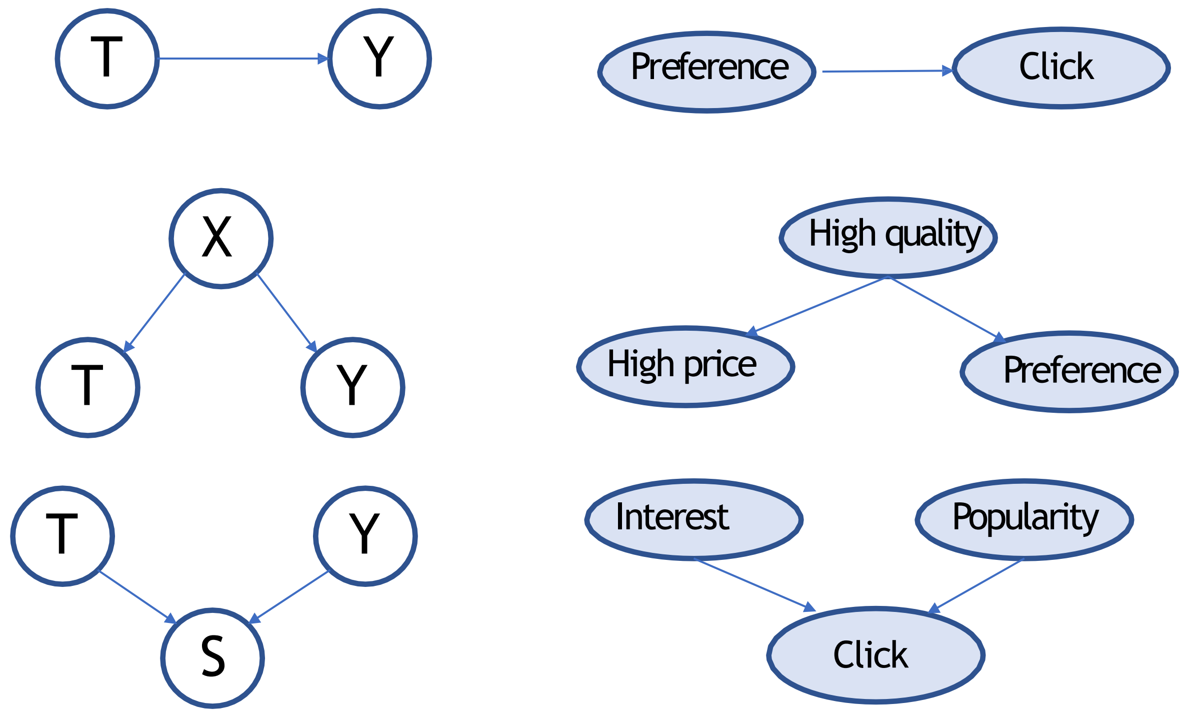
- Bias in data (Collecting):
  - Data is observational rather than experimental (missing-not-at-random)
  - Affected by many hidden factors:
    - Public opinions

    … …
- Drift along time:
  - User/item feature changes
    - Income, marriage status
    - iPhone 12 (2 0 2 1 → 2 0 2 2)
  - Preference evolution



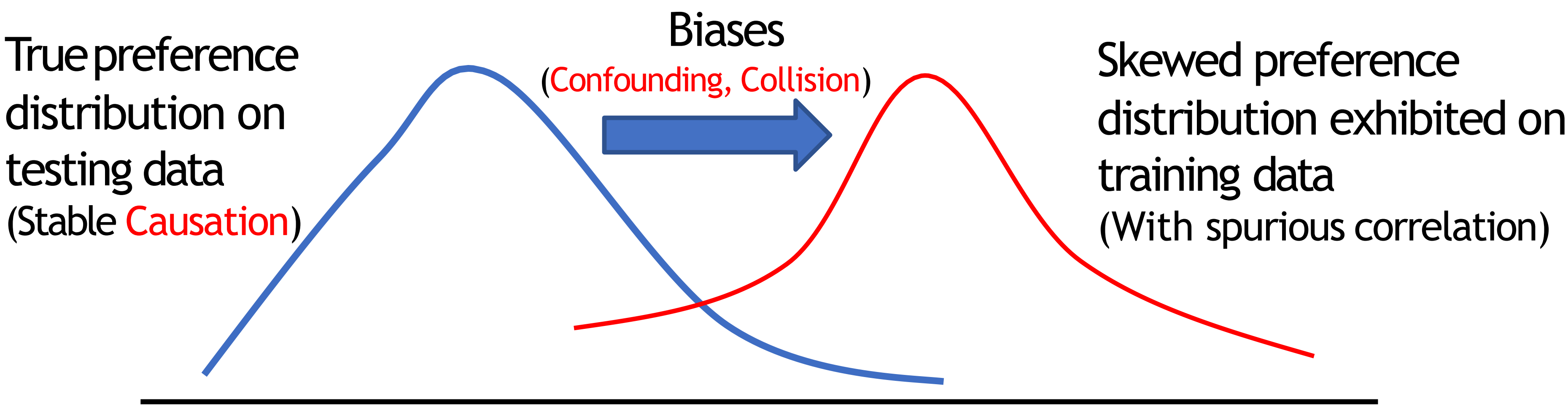
Shortcomings of Data-Driven Methods

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



Shortcomings of Data-Driven Methods

- Data-driven methods would learn skewed user preference:



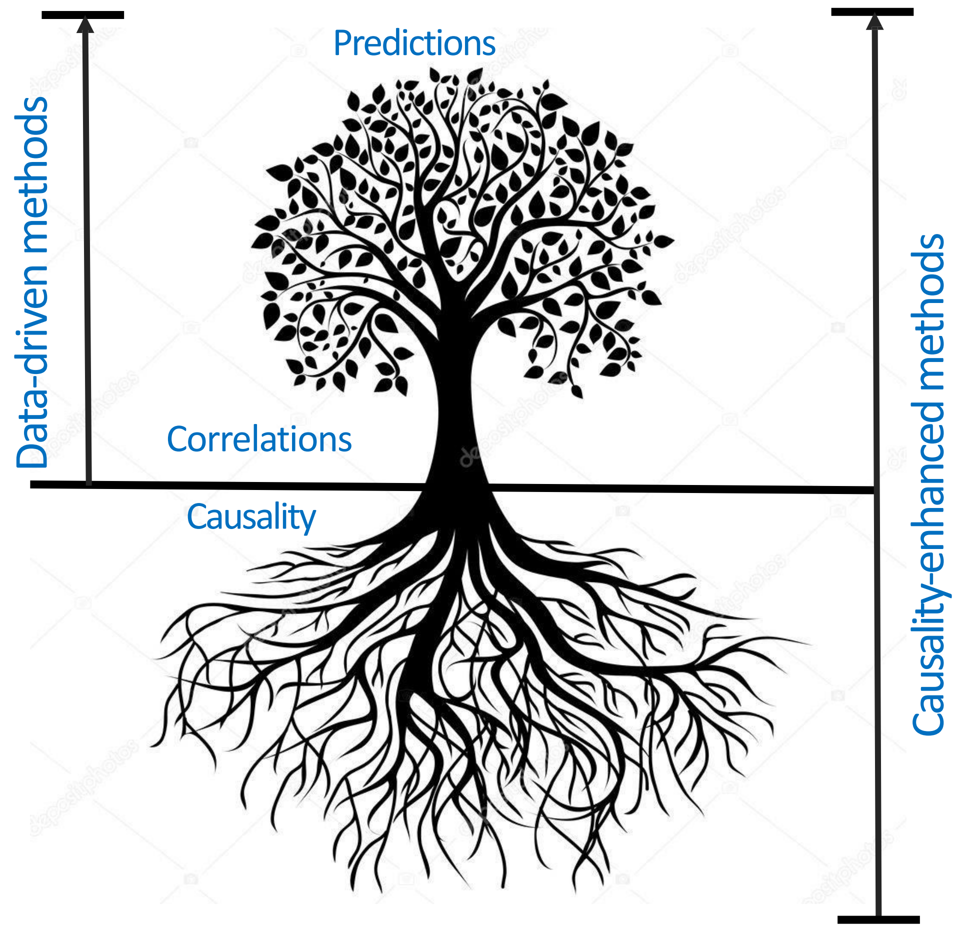
- Data-driven methods may infer spurious correlations, which are deviated from reflecting user true preference and lack interpretation.

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

- **Why Causal Inference?**

- Aim: Understanding the inherent causal mechanism of user behavior
  - Capturing user true preference

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



Task: Estimating Rating Prediction Accuracy

- Many ML methods can be applied for this task
- Standard evaluation measure:

$$R(\hat{Y}) = \frac{1}{U \cdot I} \sum_{u=1}^{U} \sum_{i=1}^{I} \delta_{u,i}(Y, \hat{Y})$$

U: # of users; I: # of items

Task: Estimating Rating Prediction Accuracy

- Many ML methods can be applied for this task
- Standard evaluation measure:

$$R(\hat{Y}) \quad = \quad \frac{1}{U \cdot I} \sum_{u=1}^{U} \sum_{i=1}^{I} \delta_{u,i}(Y, \hat{Y})$$

δ: prediction error

$$\begin{aligned}
\text{MAE:} \qquad & \delta_{u,i}(Y, \hat{Y}) = |Y_{u,i} - \hat{Y}_{u,i}| \;\; , \\
\text{MSE:} \qquad & \delta_{u,i}(Y, \hat{Y}) = (Y_{u,i} - \hat{Y}_{u,i})^2 \;\; , \\
\text{Accuracy:} \qquad & \delta_{u,i}(Y, \hat{Y}) = \mathbf{1}\{\hat{Y}_{u,i} = Y_{u,i}\} \;\; .
\end{aligned}$$

Task: Estimating Rating Prediction Accuracy

- Many ML methods can be applied for this task
- Standard evaluation measure:

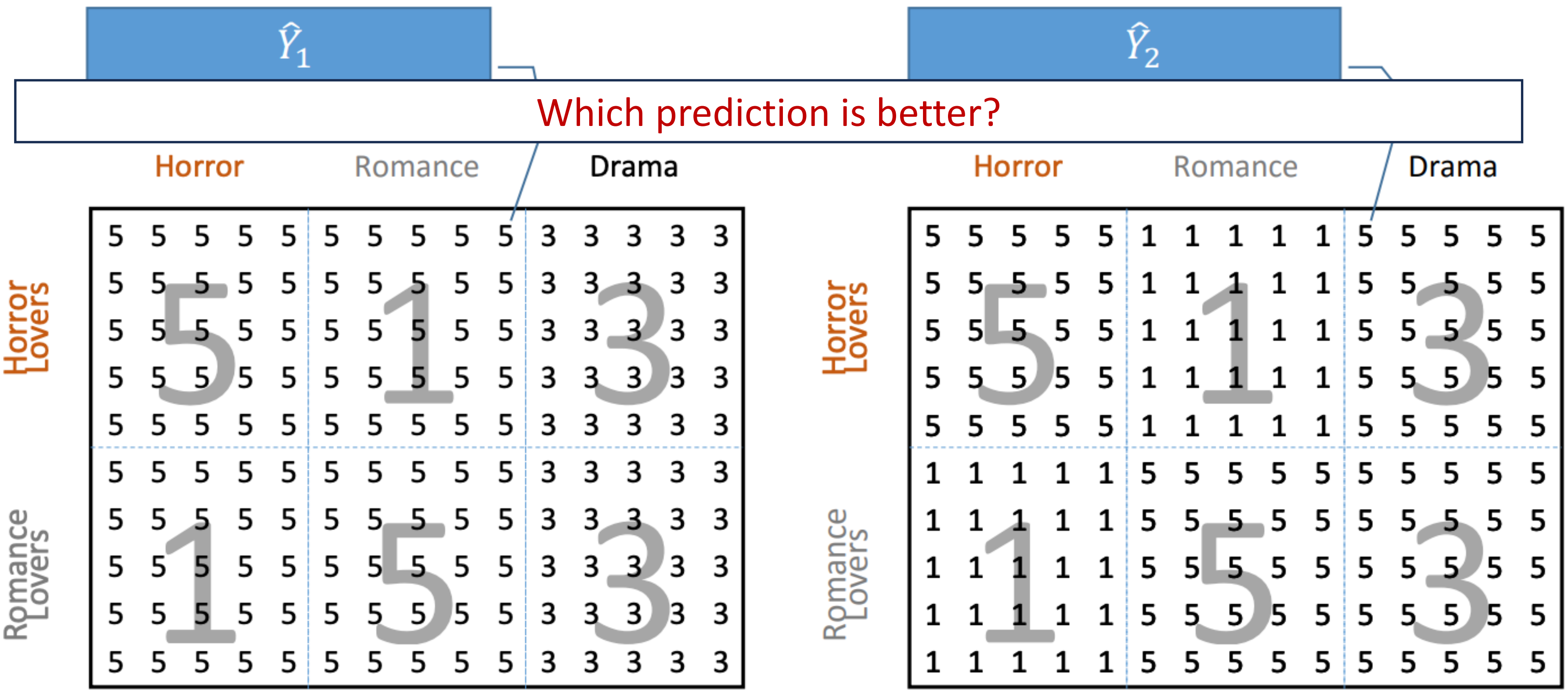
$$R(\hat{Y}) \quad = \quad \frac{1}{U \cdot I} \sum_{u=1}^{U} \sum_{i=1}^{I} \delta_{u,i}(Y, \hat{Y})$$

- Naïve estimator: the conventional way -- estimate $R(\hat{Y})$ with average over only the observed entries

$$\hat{R}_{naive}(\hat{Y}) \;=\; \frac{1}{|\{(u,i) : O_{u,i} = 1\}|} \sum_{(u,i):O_{u,i}=1} \delta_{u,i}(Y, \hat{Y})$$

$O_{u,i} = 1$ means rating (u, i) is observed

Revisit: Evaluating recommendations under Selection Bias



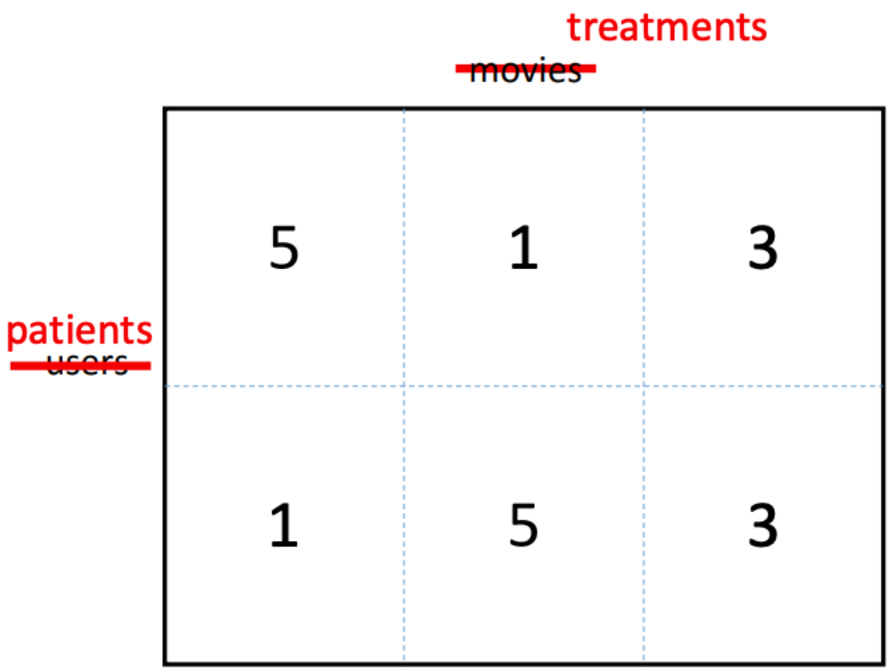
Due to selection bias, native estimator will claim $\hat{Y}_1$ better than $\hat{Y}_2$, even though $\hat{Y}_2$ is clearly better

Recommendations as Treatments

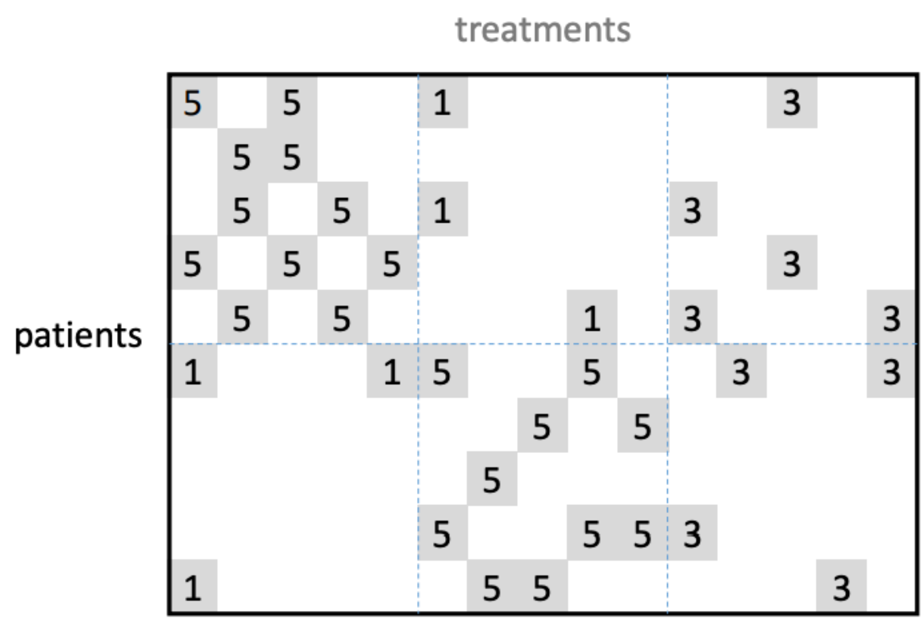
Question: How can we fix the effects of selection bias?

- Connection to potential outcomes framework

Counterfactual Outcomes Y



Observed Outcomes $\tilde{Y}$



⇒ Understand **assignment mechanism** (Imbens & Ruben, 2015)

Debiasing Evaluation

- **Assignment mechanism** for recommendation:
  - $P_{u,i} = P(O_{u,i} = 1)$

Propensities P

Horror	Romance	Drama
p	p/10	p/2
p/10	p	p/2

Debiasing Evaluation

- **Assignment mechanism** for recommendation:
  - $P_{u,i} = P(O_{u,i} = 1)$

Propensities P

Horror	Romance	Drama
p	p/10	p/2
p/10	p	p/2

Use **Inverse-Propensity-Scoring Estimator (IPS)** to obtain unbiased estimate:

$$\hat{R}_{IPS}(\hat{Y}|P) = \frac{1}{U \cdot I} \sum_{(u,i):O_{ui}=1} \frac{1}{P_{u,i}} \left(Y_{u,i} - \hat{Y}_{u,i}\right)^2$$

(Little & Rubin, 2002; Cortes et al., 2008; Bickel et al., 2009; Sugiyama & Kawanabe, 2012).

Schnabel, Tobias, et al. "Recommendations as treatments: Debiasing learning and evaluation." *ICML*, 2016.

Debiasing Learning

- Empirical Risk Minimization (ERM) successful in many settings (Cortes & Vapnik, 1995)
- Use ERM together with Inverse-Propensity-Scoring Estimator (IPS)

$$\hat{Y}^{ERM} = \underset{\hat{Y} \in \mathcal{H}}{\operatorname{argmin}} \left\{ \hat{R}_{IPS}\left(\hat{Y} \mid P\right) \right\}$$

Schnabel, Tobias, et al. "Recommendations as treatments: Debiasing learning and evaluation." *ICML*, 2016.

Debiasing Learning

- Empirical Risk Minimization (ERM) successful in many settings (Cortes & Vapnik, 1995)
- Use ERM together with Inverse-Propensity-Scoring Estimator (IPS)

$$\hat{Y}^{ERM} = \underset{\hat{Y} \in \mathcal{H}}{\operatorname{argmin}} \left\{ \hat{R}_{IPS}\left(\hat{Y} \mid P\right) \right\}$$

- For matrix factorization with mean squared error (MSE) loss:

$$\hat{Y}^{ERM} = \underset{V,W}{\operatorname{argmin}} \left\{ \sum_{O_{u,i}=1} \underbrace{\frac{1}{P_{u,i}}}_{\text{propensity weight}} \left(Y_{u,i} - V_u W_i\right)^2 + \lambda\left(\|V\|_F^2 + \|W\|_F^2\right) \right\}$$

Schnabel, Tobias, et al. "Recommendations as treatments: Debiasing learning and evaluation." *ICML*, 2016.

Propensity Estimation

Two settings:

- **Experimental** – Propensities are under our control; known by design (e.g., ad placement)
- **Observational** – Users self-select; need to estimate $P_{u,i}$
  - Estimate parameter of **binary** random variables:

$$P_{u,i} = P(O_{u,i} = 1 \mid X, \tilde{Y})$$

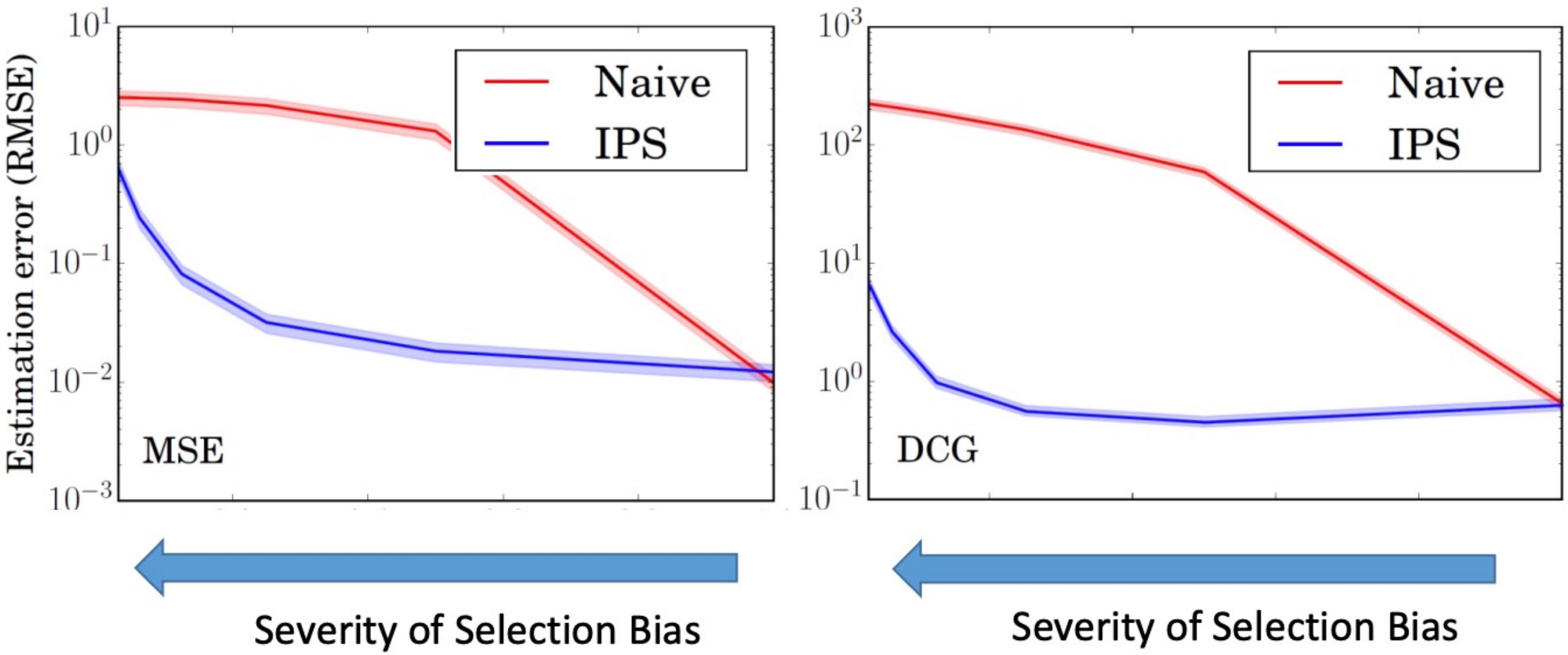
  - **Variety of models:** Logistic Regression, Naïve Bayes, etc.

Observations O

Horror					Romance					Drama				
1	0	1	0	0	1	0	0	0	0	0	0	1	0	0
0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	1	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	1	0	1	0	0	0	0	1	0	1	0	0	0	1
1	0	0	0	1	1	0	0	1	0	0	1	0	0	1
0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
0	0	0	0	0	1	0	0	0	1	1	0	0	0	0
1	0	0	0	0	0	1	1	0	0	0	0	0	1	0

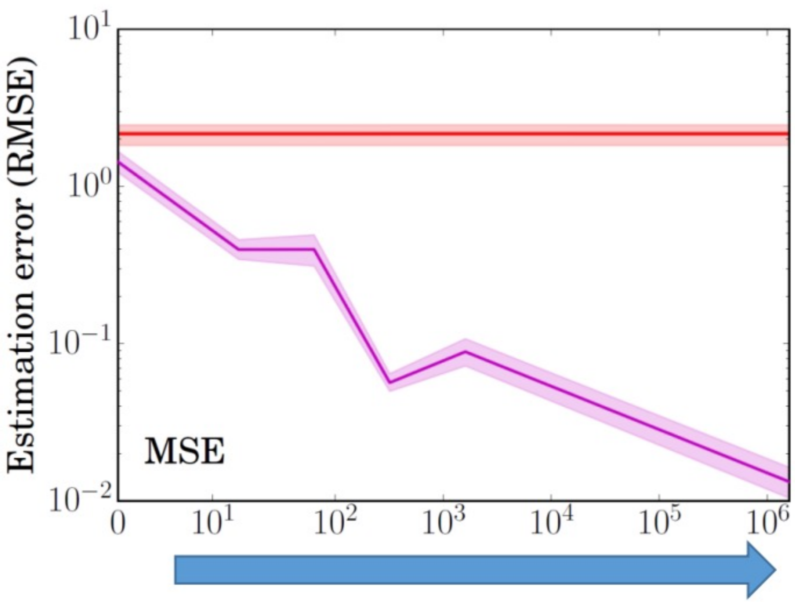
Debiasing Evaluation

- Robustness to selection bias

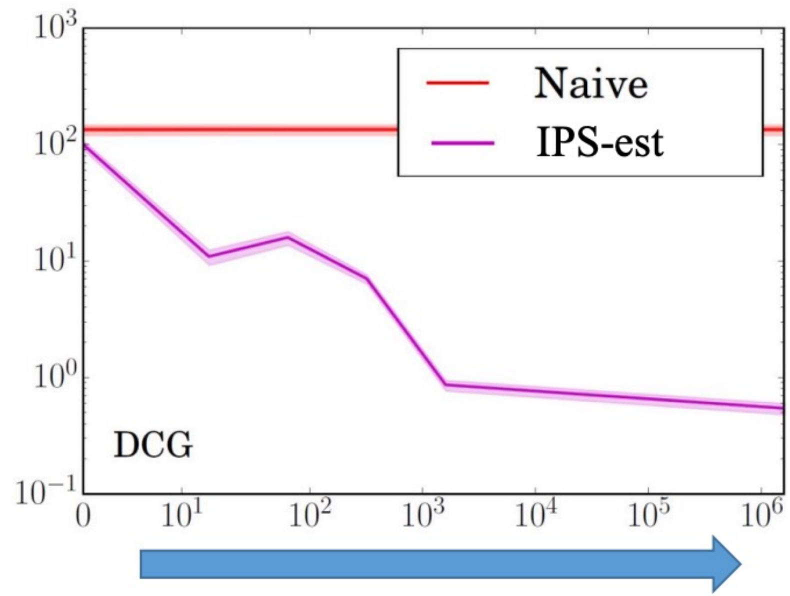


Debiasing Evaluation

- Robustness to inaccurate propensities:



More accurate propensities



More accurate propensities

Outline

- Causal inference in recommender system
  - Recommendation as treatment
  - Biases in recommendation: a causal view
- Other directions

Biases in RS

- The introduction of causal techniques into recommender systems (RS) has brought great development to this field and has gradually become a trend.

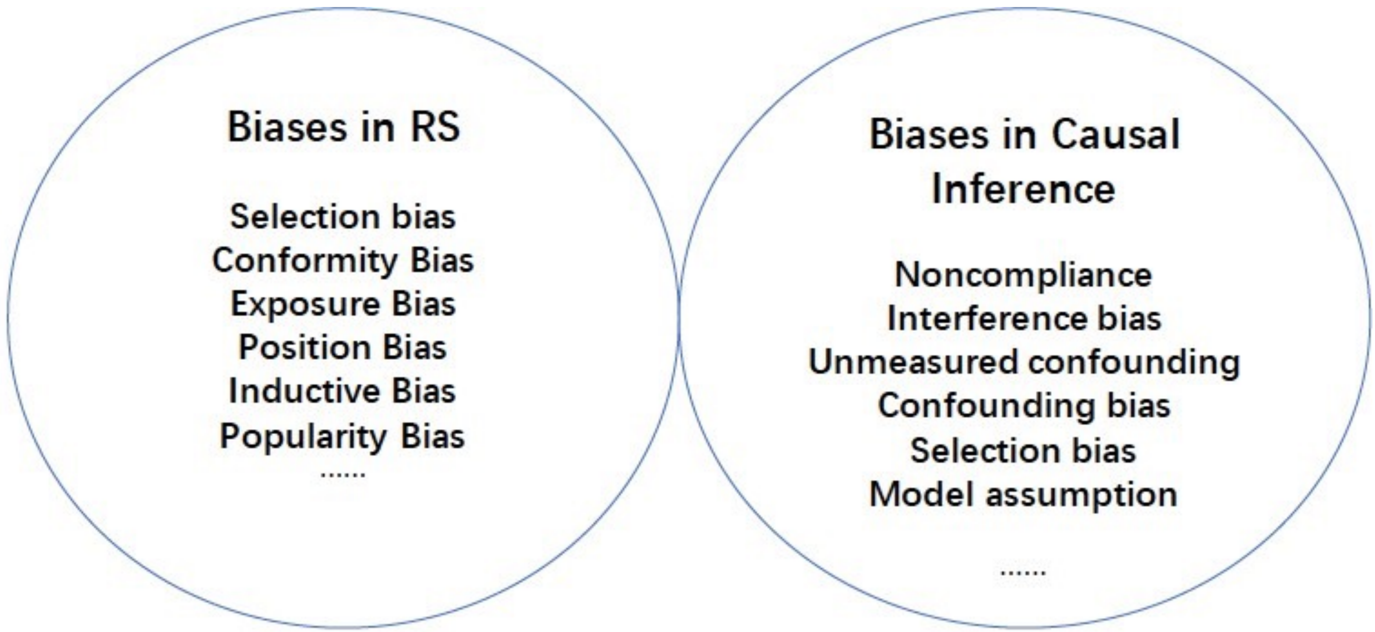
- various biases exist in observed data. Yet, formal definitions of the biases in RS are still not clear, which leads to difficulty in discussing theoretical properties and limitations of various debiasing approaches.

- This greatly hinder the development of RS.

Jiawei Chen and Hande Dong and Xiang Wang and Fuli Feng and Meng Wang and Xiangnan He (2020), ' Bias and Debias in Recommender System: A Survey and Future Directions', arXiv:2010.03240.

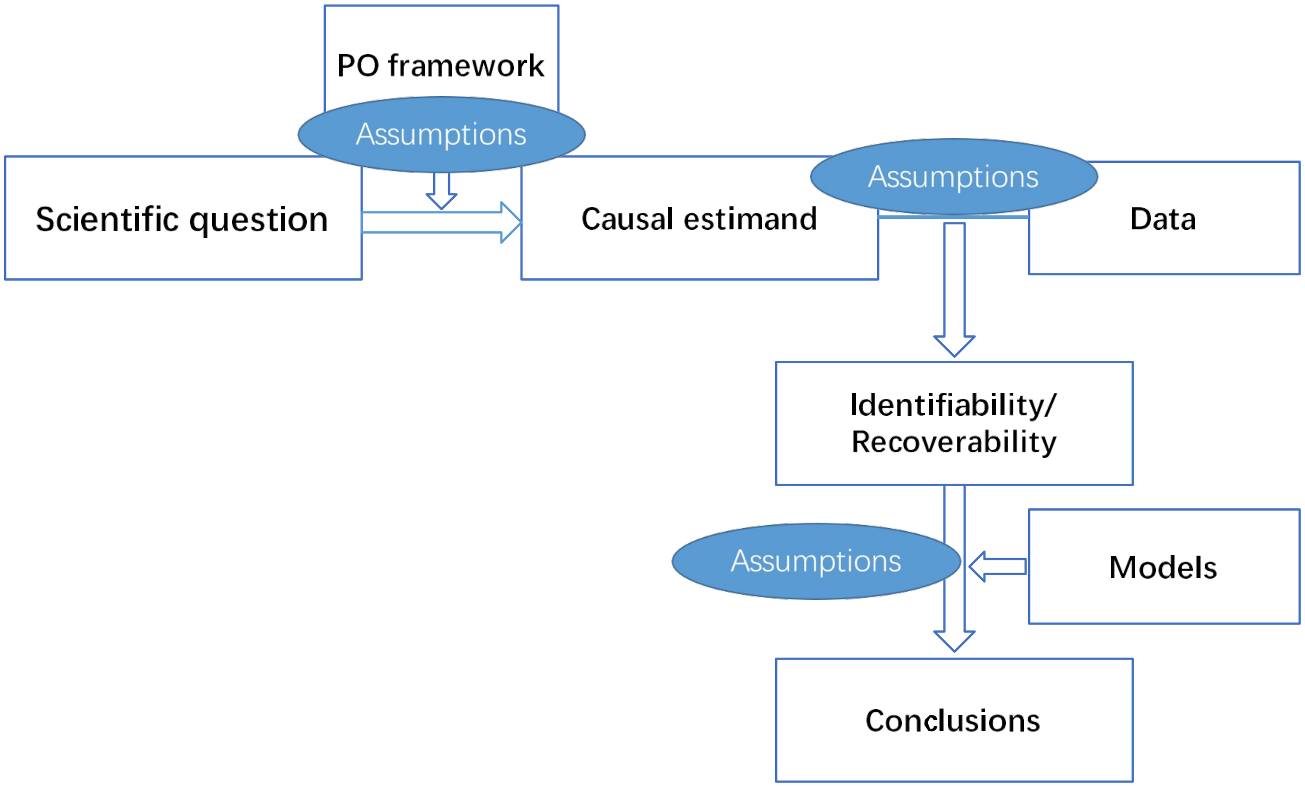
Peng Wu, Haoxuan Li, Yuhao Deng, Wenjie Hu, Quanyu Dai, Zhenhua Dong, Jie Sun, Rui Zhang, Xiao-Hua Zhou (2021), ``Causal Analysis Framework for Recommendation'', arXiv:2201.06716. (To appear in IJ-CAI)

Biases in RS: A Causal View



- Provide formal definitions of various biases in RS.

Biases in Causal Inference



We need a variety of assumptions to climb from association (data) to causality (causal conclusions), violating these assumptions may result in various biases.

New perspective of biases in RS.

	Assumptions	Biases in causal inference	Biases in RS
Define causal estimands	SUTVA(a)	undefined	position bias
	SUTVA(b)	interference bias	conformity bias
Recoverability	consistency	noncompliance	undefined
	positivity	undefined	exposure bias
	exchangeability	confounding bias	popularity bias
	conditional exchangeability	hidden confounding bias	undefined
	random sampling	selection bias	user selection bias, exposure bias
Model	model specification	model mis-specification	inductive bias

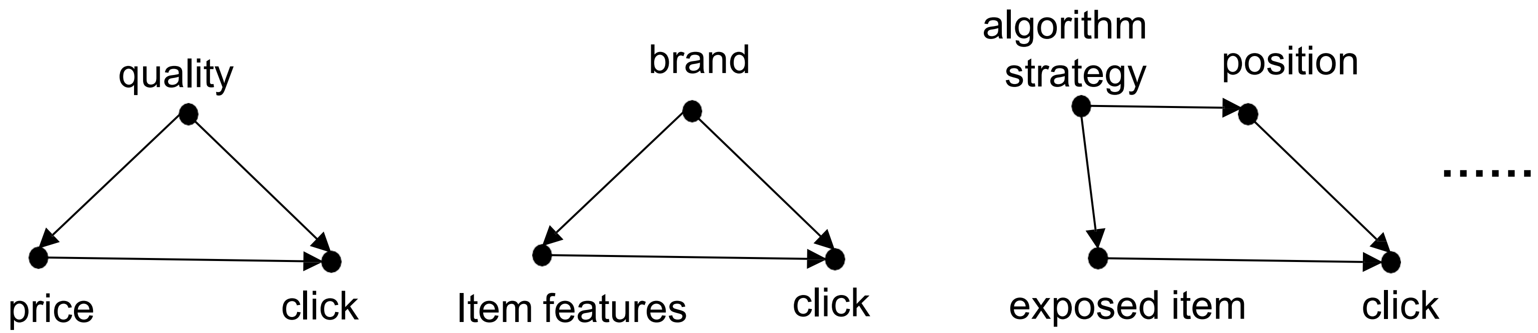
- According to this table, we can define the descriptive biases in RS formally using the rigorous syntax of causal inference.
- It also provides an opportunity to apply the existing causal inference methods to RS.
- In addition, for the unique characteristics of RS, we expect that a series of new methods will be developed by weakening or substituting the assumptions.

Jiawei Chen and Hande Dong and Xiang Wang and Fuli Feng and Meng Wang and Xiangnan He (2020), ' Bias and Debias in Recommender System: A Survey and Future Directions', arXiv:2010.03240.

Peng Wu, Haoxuan Li, Yuhao Deng, Wenjie Hu, Quanyu Dai, Zhenhua Dong, Jie Sun, Rui Zhang, Xiao-Hua Zhou (2021), ``Causal Analysis Framework for Recommendation'', arXiv:2201.06716. (To appear in IJ-CAI)

Confounding in recommendation

- Are there confounders in recommendation?
  - There are some possible examples



  - What's more, some confounder are observable/measurable, some confounder are unobservable/unmeasurable.
    e.g., company is measurable, quality is unmeasurable.

Existing work regarding observed confounders

- Existing work

The backdoor adjustment is obvious selection, and most work is based on it.

2021 SIGIR&KDD

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

2021CIKM

- Gupta et.al. CauSeR

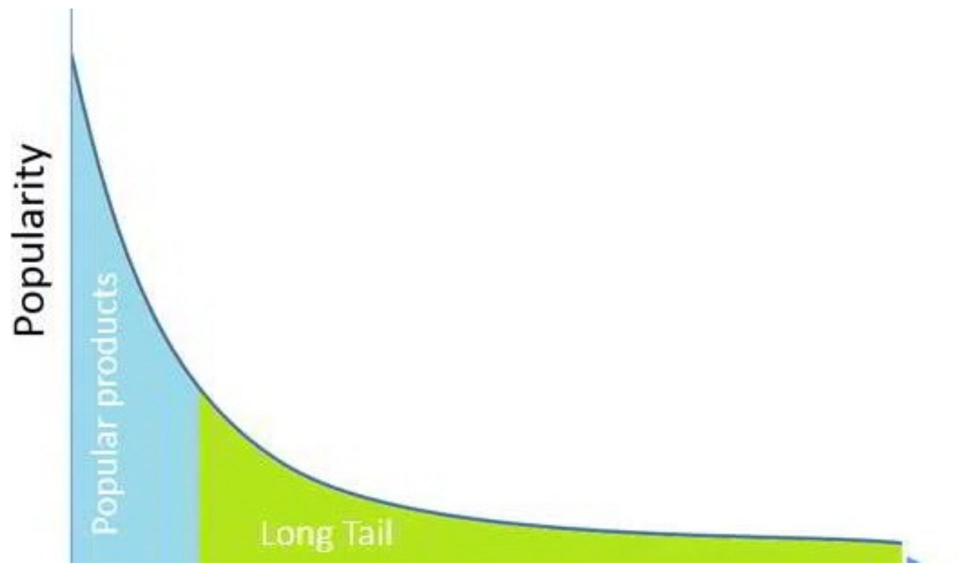
2022

- Zhang et.al. DCR
- Wang et.al. CaDSI

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

PDA: Confounding view of the popularity bias

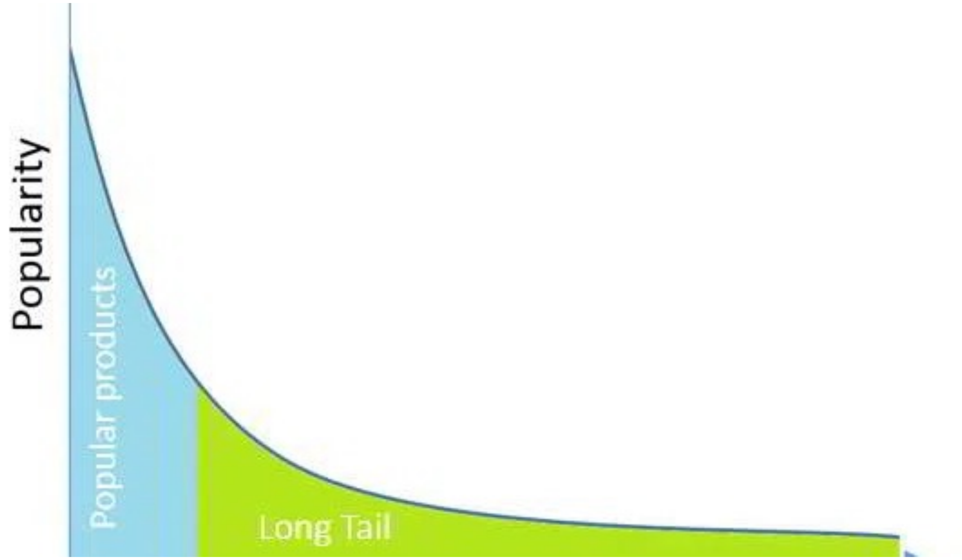
- Popularity bias
  - Favor a few popular items while not giving deserved attention to the majority of others
  - The popular items are recommended even more frequently than their popularity would warrant, amplifying long-tail effects.



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

PDA: Confounding view of the popularity bias

- 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).

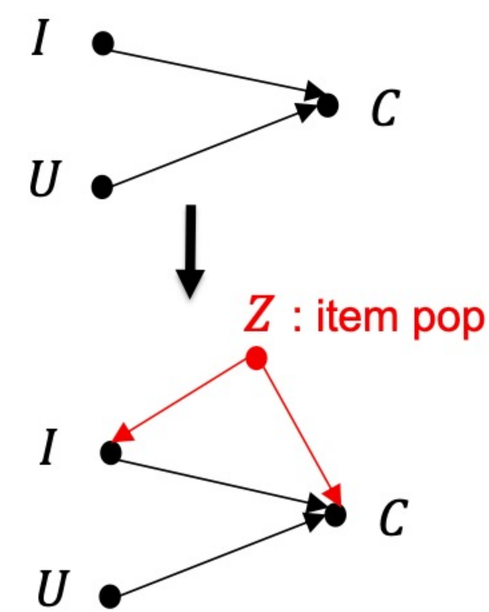


Zhang et al. SIGIR 2021. Causal Intervention for Leveraging Popularity Bias in Recommendation

PDA: Confounding view of the popularity bias

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

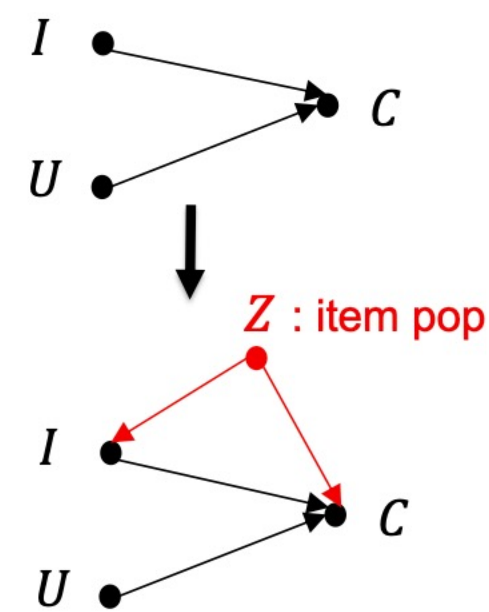
U: user**; I: exposed item;**
C: interaction label



PDA: Confounding view of the popularity bias

- What is the bad effect of popularity bias?
  - Common 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.

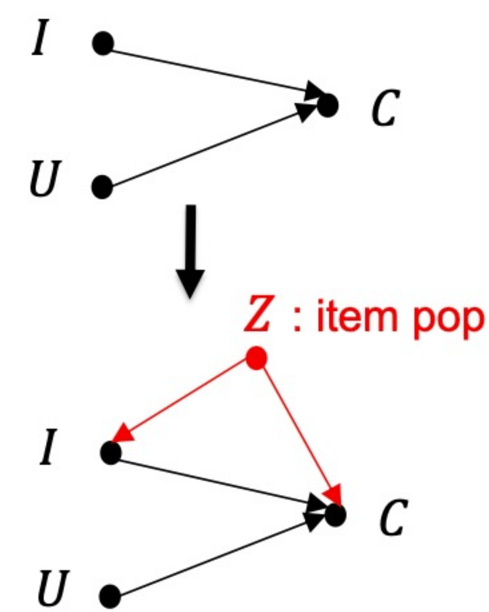
U: user; **I: exposed item;**
C: interaction label



PDA: Confounding view of the popularity bias

- What is the bad effect of popularity bias?
  - Common 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.

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



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

Correlation:

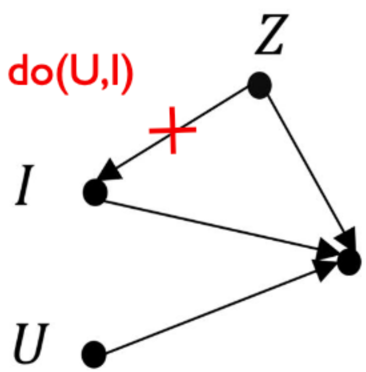
Bad effect

$$P(C|U,I) = \sum_Z P(C|U,I,Z)\boxed{P(Z|I)}$$
$$\propto \sum_Z P(C|U,I,Z)\boxed{P(I|Z)}P(Z)$$

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

PDA: Confounding view of the popularity bias

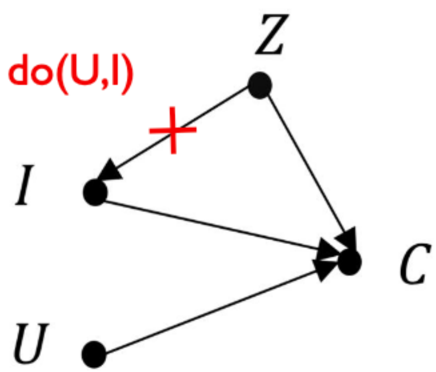
- Training & Inference: Popularity De-confounding (PD, remove bad effect)



- To estimate $$P(C|do(U,I)) = \sum_{z} P(C|U,I,z)P(z)$$

PDA: Confounding view of the popularity bias

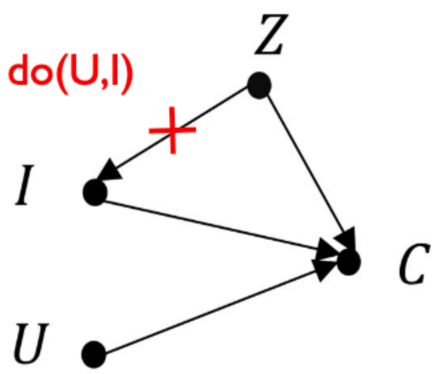
- Training & Inference: Popularity De-confounding (PD, remove bad effect)



- 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

PDA: Confounding view of the popularity bias

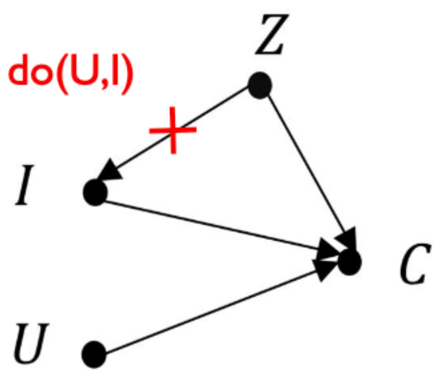
- Training & Inference: Popularity De-confounding (PD, remove bad effect)



- 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 $P(\mathrm{C|do(U,I)})$
    - $\sum_Z P(C|U,I,Z)P(Z) \propto f_\Theta(u,i)$
    - Derivation sees the paper

PDA: Confounding view of the 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

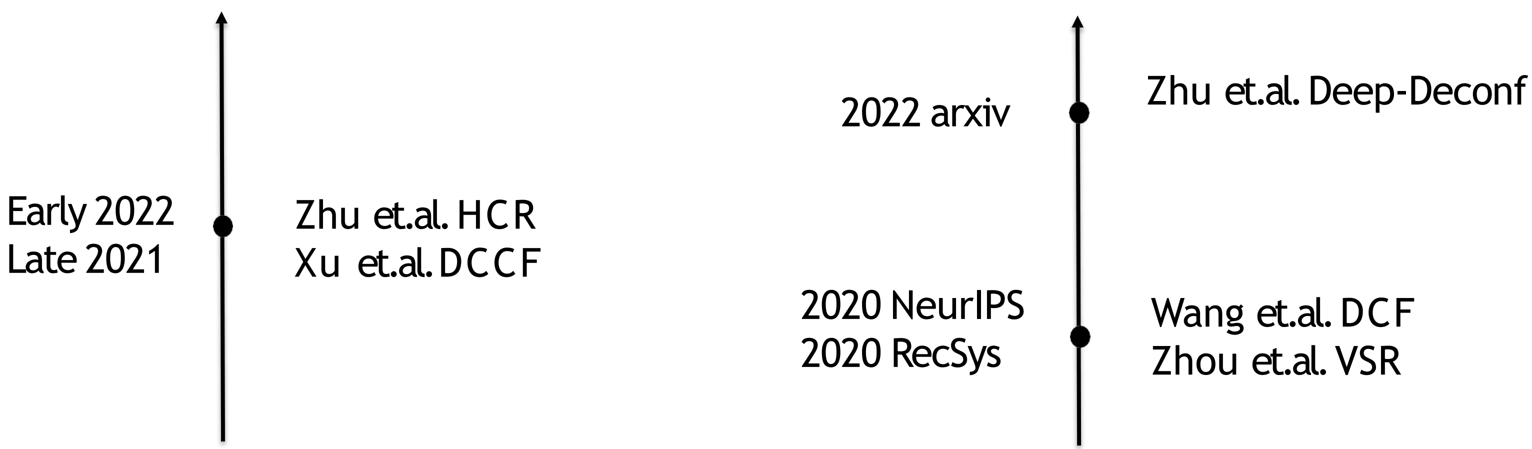
- Another Inference: Popularity Adjusting (inject desired popularity bias)
  - Inject the desired pop bias $\tilde{Z}$ by causal intervention

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

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

Existing Work 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:

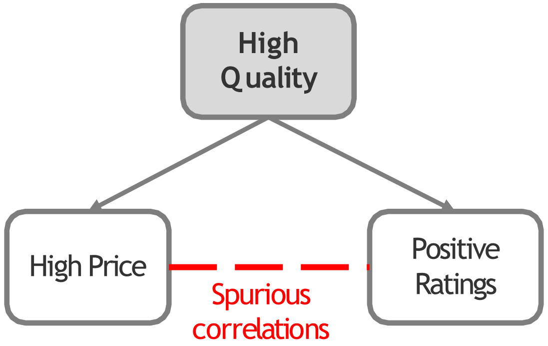


Front-door adjustment

Learning substitutes

HCR: The Front-door Adjustment-based Method

- Some confounders are hard to measure.
  - Technical difficulties, privacy restrictions, *etc.*
  - E.g., product quality.

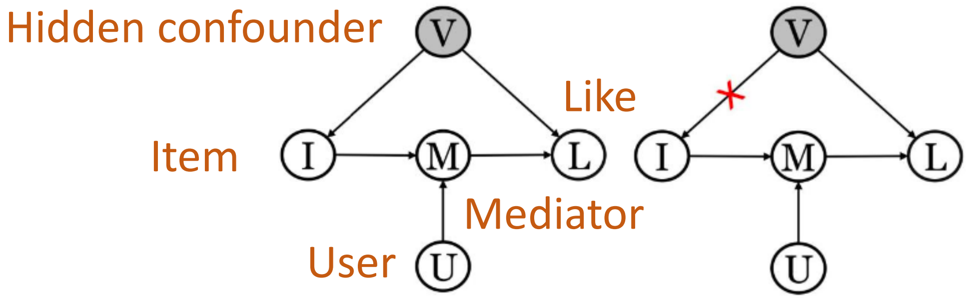


- Removing hidden confounders is hard:
  - Inverse Propensity Weighting
    - Based on strict assumption of no hidden confounder.
  - Backdoor Adjustment
    - Require the confounder's distribution.

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

HCR: The Front-door Adjustment-based Method

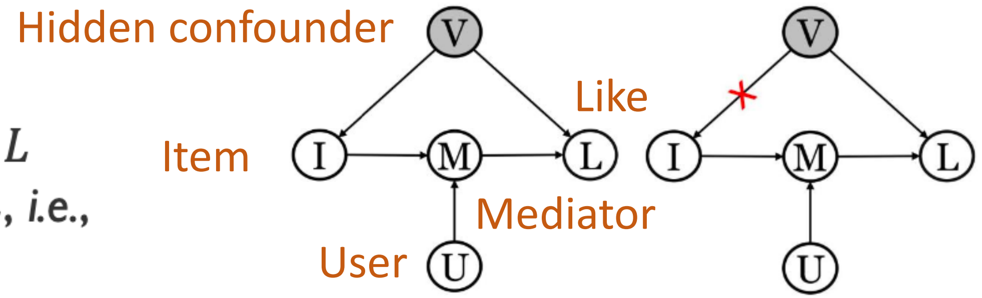
- Abstract user feedback generation process into causal graph.
  - V: hidden confounder; L: like feedback; I: item; U: user.
  - M: a set of variables that act as mediators between $\{U, I\}$ and L, e.g., user-item feature matching, and click.



HCR: The Front-door Adjustment-based Method

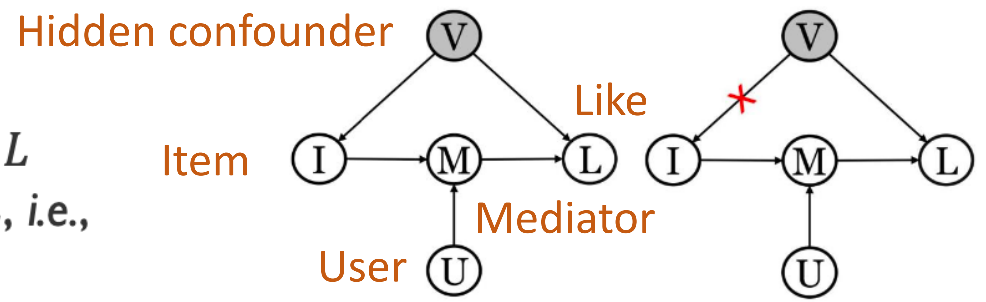
- Abstract user feedback generation process into causal graph.
  - V: hidden confounder; L: like feedback; I: item; U: user.
  - M: a set of variables that act as mediators between $\{U, I\}$ and L, e.g., user-item feature matching, and click.
- Key:
  - Block the backdoor path $I \leftarrow V \rightarrow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

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



HCR: The Front-door Adjustment-based Method

- 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.,

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

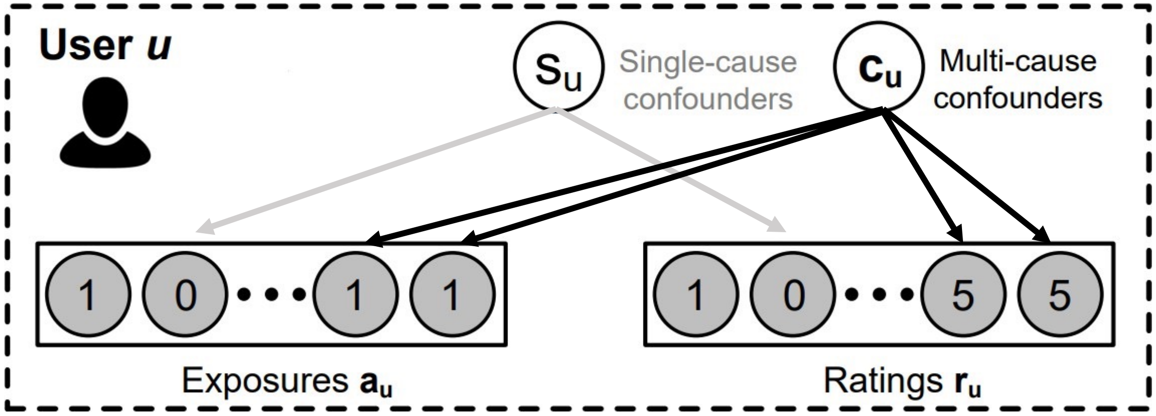
- Hidden Confounder Removal (HCR) framework.
  - Front-door adjustment
    - decompose causal effect of I on L into: 1) the effects of I on M and 2) the effect of M on L.

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

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

Learning Substitutes-based Method

- 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.



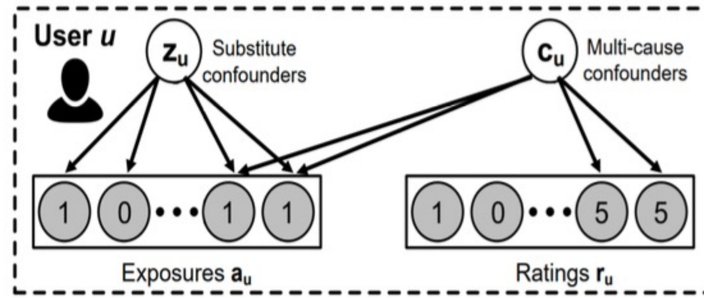
Wang et al. RecSys 2020. Causal inference for recommender system.
Zhu et.al. Arxiv 2022. Deep causal reasoning for recommendations.

Learning Substitutes-based Method

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

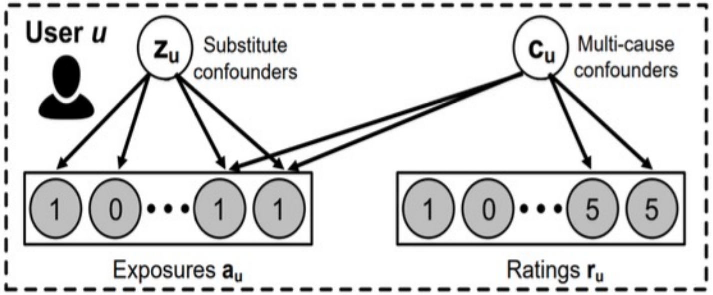


Learning Substitutes-based Method

- 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

-

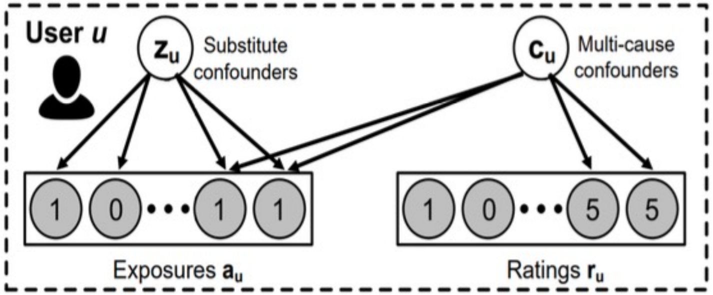
Wang et al. RecSys 2020. Causal inference for recommender system.
Wang et.al. *J Am Stat Assoc 2019.* The blessings of multiple causes.
Zhu et.al. Arxiv 2022. Deep causal reasoning for recommendations.

Learning Substitutes-based Method

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



- Step I: learning substitutes

  Finding a Z_u, such that:
  $p(a_{u1}, \ldots, a_{um}|z_u) = \prod_i p(a_{ui}|z_u)$

  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

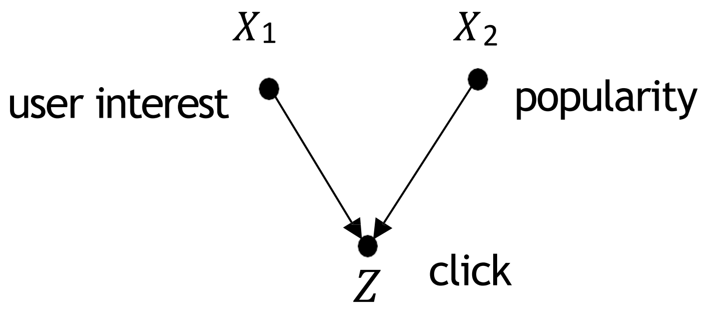
  Control the substitutes to fit recommender model

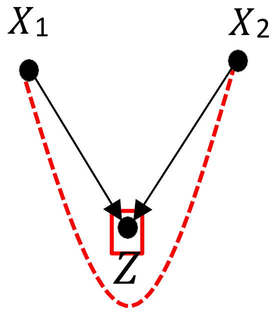
  Example:
  $y_{ui}(a) = \theta_u^\top \beta^i \cdot a + \gamma_u \cdot z_{ui} + \epsilon_{ui}$
  where θ_u and β_i refer user preference and item attributes, respectively.

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.

Colliders in Recommendation

- Are there colliders in recommendation?
  - E.g., clicking is affected by user interest and the item popularity.

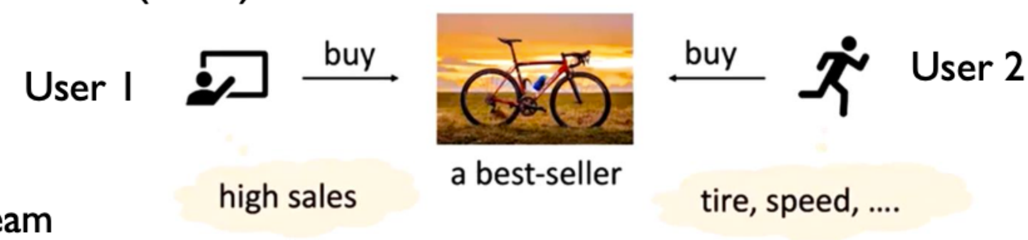




Conformity bias in RS: Colliding Effects

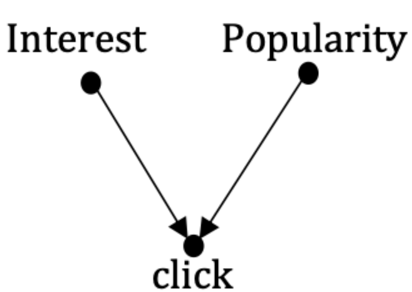
- What are causes of a user-item interaction (click)?

  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:

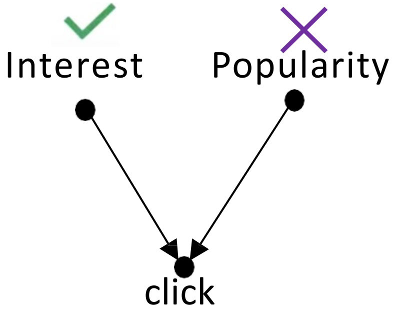


  - Interest and Popularity (conformity) are independent
  - But, they are correlated given clicks:
    A click on less popular item → High Interest

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

DICE: Colliding Effects for Disentangling True Interest

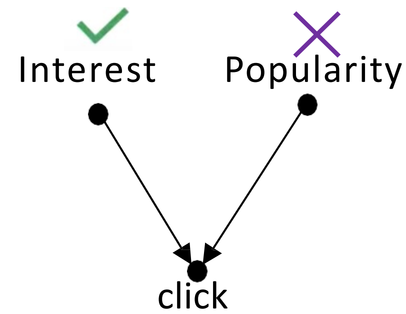
- Partial pairwise data identifies true interest:
  - O_1: {<u, pos_item, neg_item>}, wherein *pos_item* is less popular than *neg_item*}
  - Pairwise cause-specific data (interest-driven): we can ascertain that the interaction is more likely due to user



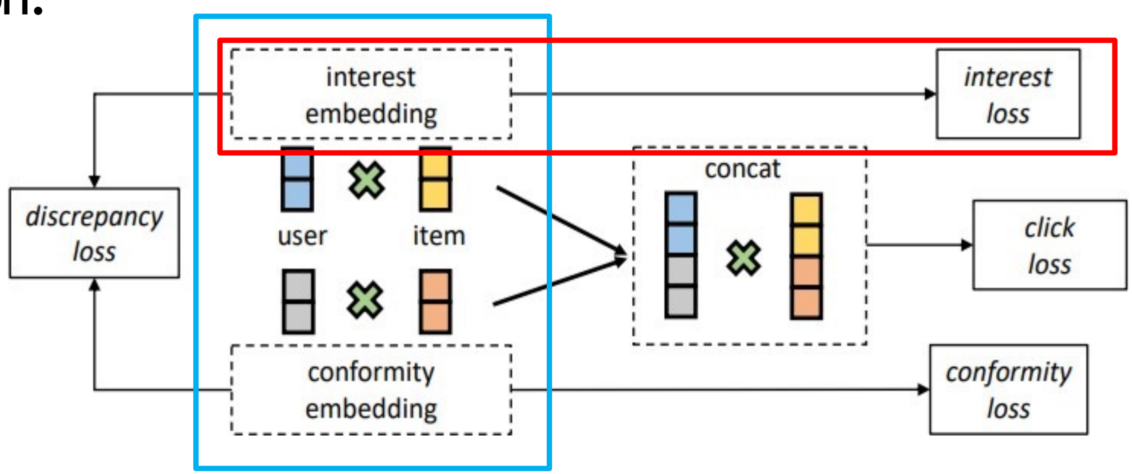
Zheng et al. WWW 2021. Disentangling User Interest and Conformity for Recommendation with Causal Embedding

DICE: Colliding Effects for Disentangling True Interest

- 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 (interest-driven): we can ascertain that the interaction is more likely due to user interest



- Solution:



- Key2: learning interest embedding on interest-driven pairwise data (O_1).

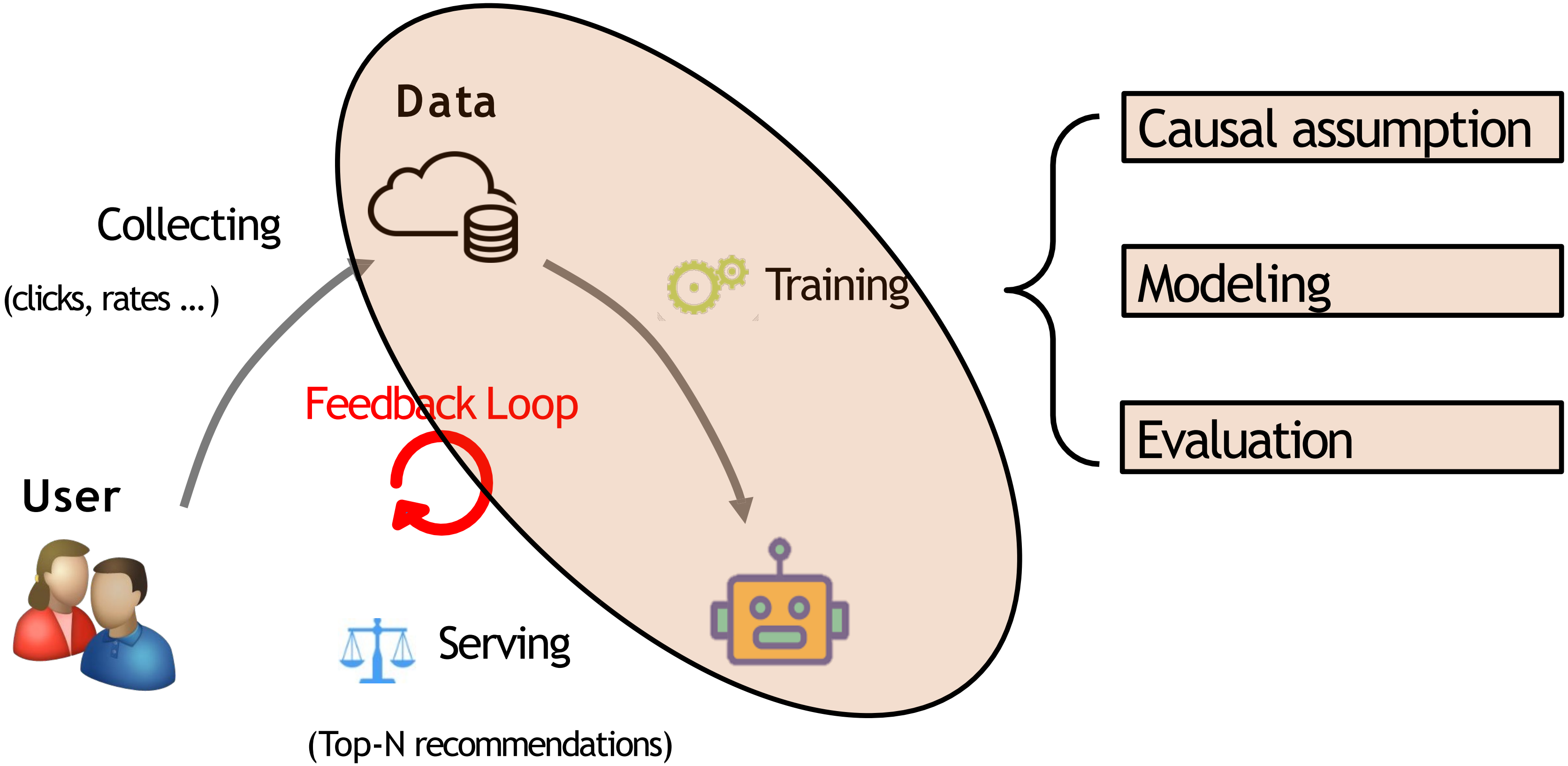
- Key1: Split user/item representation into two embeddings

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

Outline

- Causal inference in recommender system
  - Recommendation as treatment
  - Biases in recommendation: a causal view
- Other directions

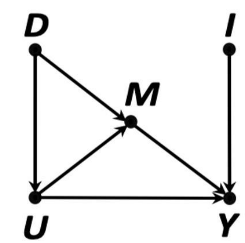
Open Problems and Future Directions



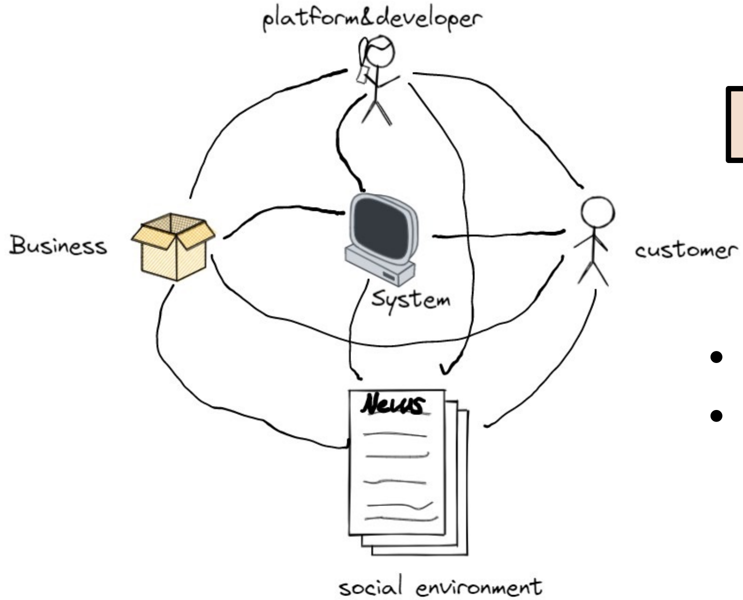
Causal Assumption

- PO & SCM requires 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

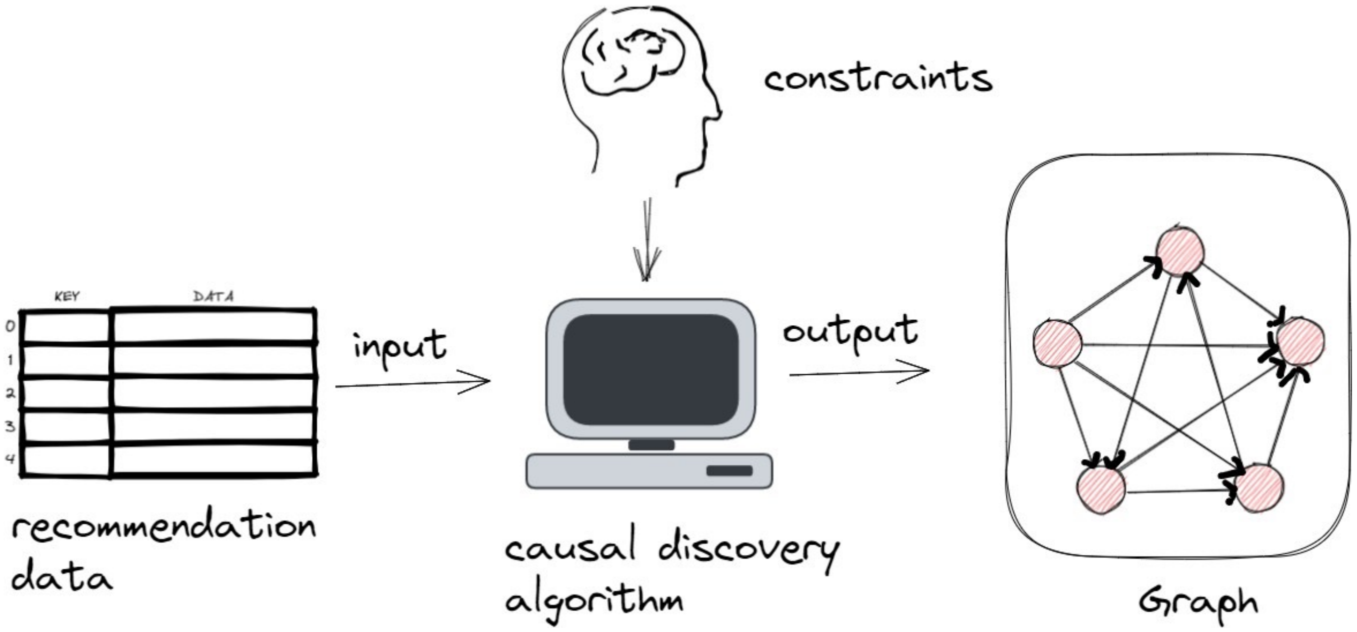


How to obtain proper causal assumptions?

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

Causal Assumption

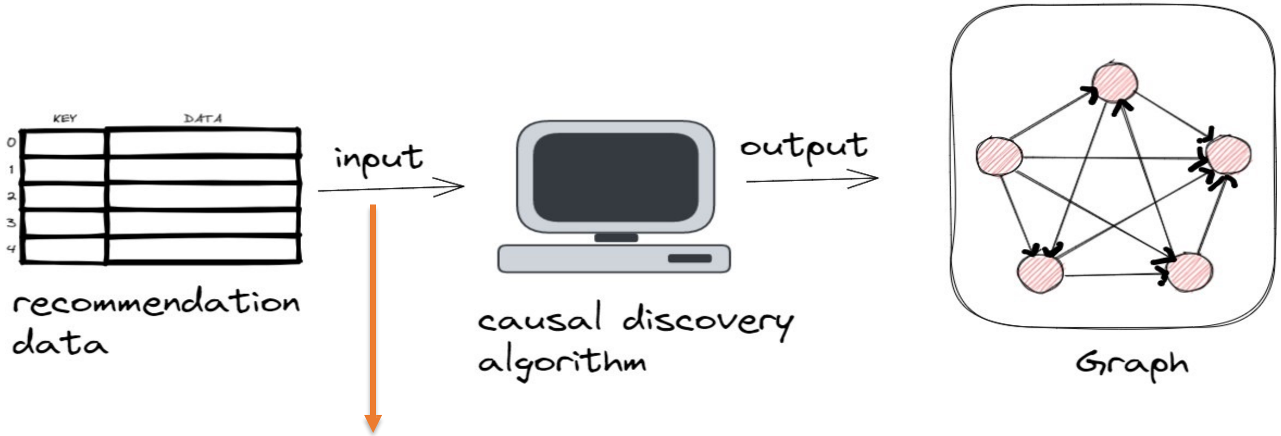
- Future direction: causal discovery in recommendation



Automatic discovery of cause graphs with causal discovery algorithms

Causal Assumption

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

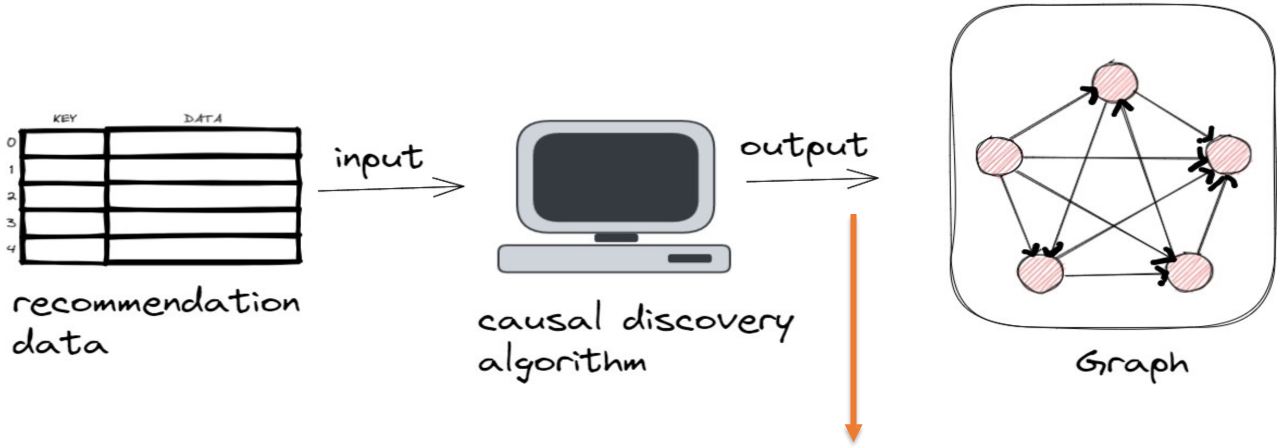


- Normal causal discovery algorithm only deals with few variables
- Challenge 1:

  High-dimensional inputs; hidden variables.

Causal Assumption

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

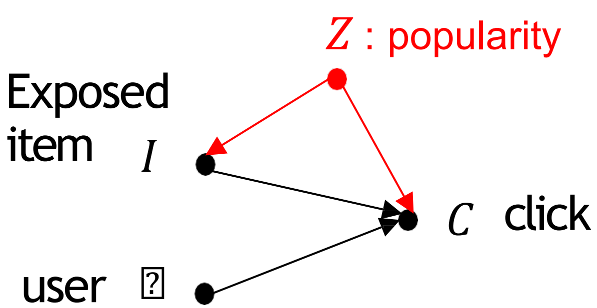


- The output usually is a set of causal graphs instead of only one graph.
- Challenge 2:

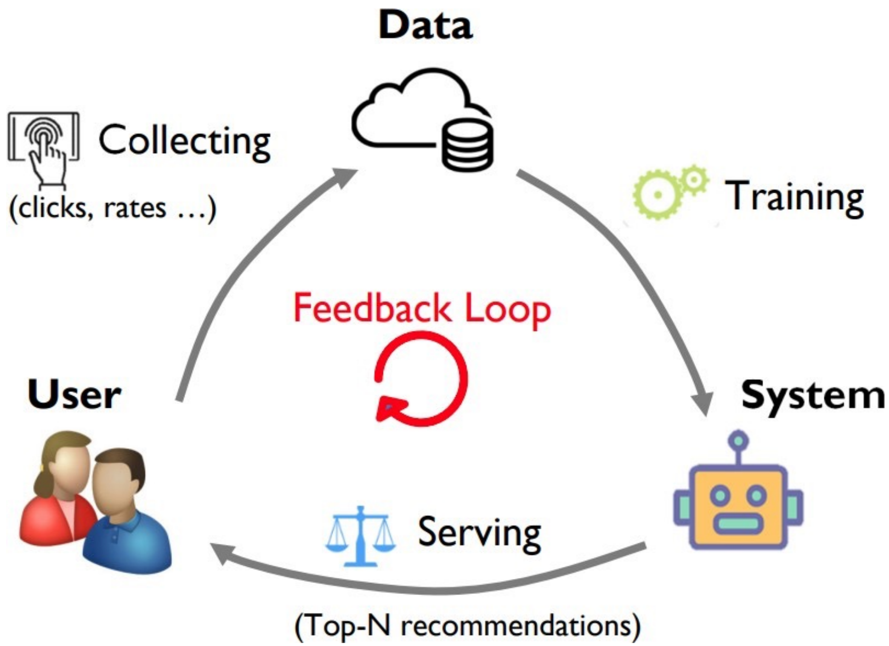
Unreliable graphs in the graph set.

Causal Modeling

- Existing work focuses on one training step



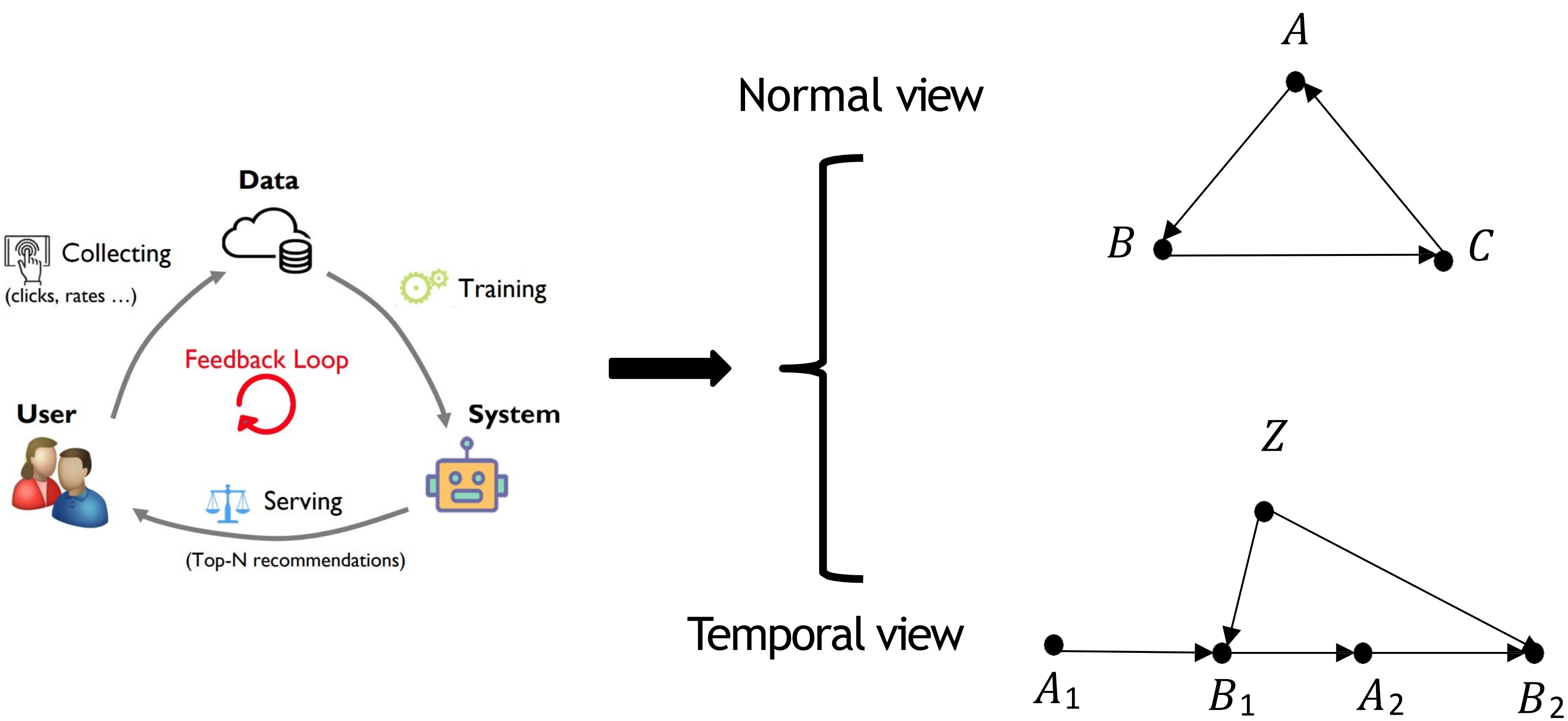
Popularity also influences the collecting step



How to model the causal effect of feedback loop?

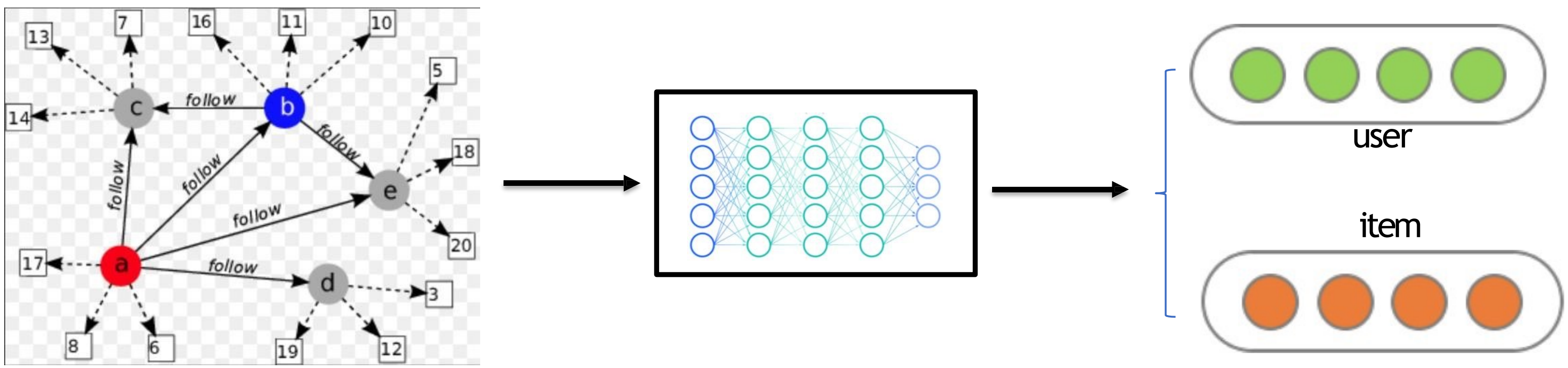
Causal Modeling

- Future direction: Temporal causal modeling



Causal Modeling

- Existing work relies on latent representation



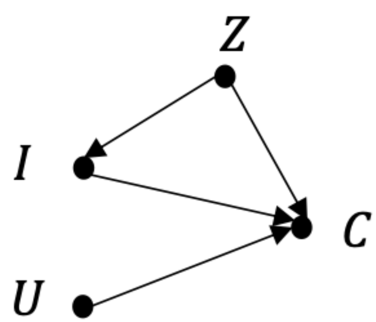
Recommendation data | Neural network | representation

- The key of many recommender models is to learn user/item representations
- But, rare work focus on injecting causation into representations

How to learn causal representation?

Causal Modeling

- **Existing work requires many manual operations**



$$\mathrm{P}(C|\mathrm{do}(U),\mathrm{do}(I)) = \sum_{z} P(C|U,I,Z)P(Z)$$

Model

$MF(u,i) \times m_i^t$

① Manually define causal assumption, e.g., casual graph

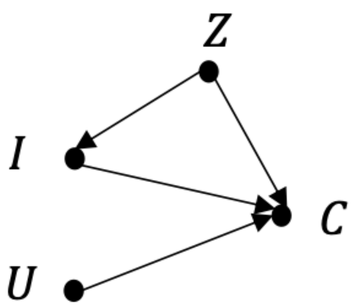
② Manually identify estimation goal according to ①

③ Manually design recommender model based on ②

How to reduce the cost of human-efforts?

Causal Modeling

- **Future direction: Auto-causal recommendation**



$$P(C|do(U),do(I)) = \sum_{z} P(C|U,I,Z)P(Z)$$

Model

$MF(u,i) \times m_i^t$

① Manually define causal assumption

Causal discovery

② Manually identify estimation goal according to ①

Query/target causal understanding + automated causal inference

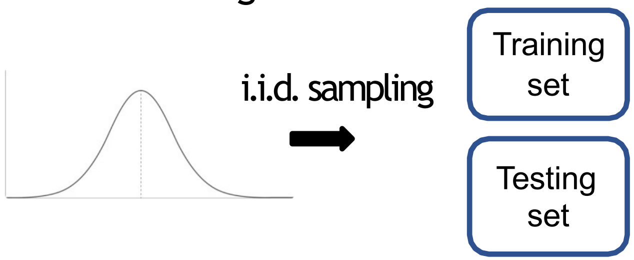
③ Manually design recommender model based on ②

Causal model automated searching

Evaluation

- One thousand papers, one thousand evaluation protocols

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



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.

Evaluation

- Future direction: causality-aware evaluation metrics

Example 1 -- the effect of recommending operation

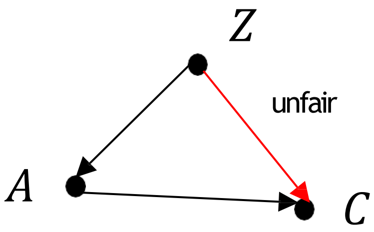
A and B are both matched to user preference, but recommending B can bring more gains.

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

Masahiro 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.



References

- Zhang, Yang, et al. "Causal intervention for leveraging popularity bias in recommendation." *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2021. (Zhang et.al. PDA)
- Wang, Wenjie, et al. "Deconfounded recommendation for alleviating bias amplification." *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2021. (wang et.al. DecSR)
- Wang, Xiangmeng, et al. "Causal Disentanglement for Semantics-Aware Intent Learning in Recommendation." IEEE Transactions on Knowledge and Data Engineering (2022). (Wang et.al. CaDSI)
- Gupta, Priyanka, et al. "CauSeR: Causal Session-based Recommendations for Handling Popularity Bias." Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021. (Gupta et.al., CauSeR)
- Yang, Xun, et al. "Deconfounded video moment retrieval with causal intervention." *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2021. (Yang et.al. DCM)
- Wang, Yixin, et al. "Causal inference for recommender systems." Fourteenth ACM Conference on Recommender Systems. 2020. (Wang et.al. DCF)

- Some slides are from:
  - Riddhiman Adib. “Drawing Causality from Observational Data in Healthcare: Approaches to Causal Modeling and Reasoning through Graphical Causal Models”.
  - Yang Zhang, et al. Causal Recommendation: Progresses and Future Directions. WWW 2022 Tutorial.

Reading Materials

- Schnabel T, Swaminathan A, Singh A, et al. Recommendations as treatments: Debiasing learning and evaluation[C]//international conference on machine learning. PMLR, 2016: 1670-1679.
- Zhang, Yang, et al. "Causal intervention for leveraging popularity bias in recommendation." *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2021. (Zhang et.al. PDA)
- Wang W, Feng F, He X, et al. Deconfounded recommendation for alleviating bias amplification[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 1717-1725.

Thank you!
Questions?