如何找新Idea?

---- 以推荐系统纠偏为例

陈佳伟
浙江大学计算机学院

Sep 2024

什么是idea?

大部分Idea：用适当方法
解决有用的问题

方法:    问题: 

怎么找idea? --- 更好的模型(借鉴)

现有方法：

观察到：

新的idea：

为何方法更
适合这个问题?

怎么找idea? --- 更好的模型(改进)

现有方法： 

发现：柄太短了

针对问题，挖掘现有方法的缺陷

新的idea： 

怎么找idea? --- 新的问题

现有方法：

新的idea：

阐明新问题的意义，分析其性质

怎么找idea? --- 分析型（方法或问题的性质）

深入研究斧头(方法)的性质

深入研究树(问题)的性质

怎么找idea? --- 怎么做?

掌握更多的方法：

关注其他领域的方法

深入理解问题：

对该领域全面调研

写好一篇文章

写好一篇文章

❖ Outline

1. **Bias in Recommendation**
    - Recommender Systems
    - Bias in Recommendation
    - The Influences of Bias
2. **My Recent Debiasing Work**

Recommender System

☐ Recommender system (RS) helps address information overload

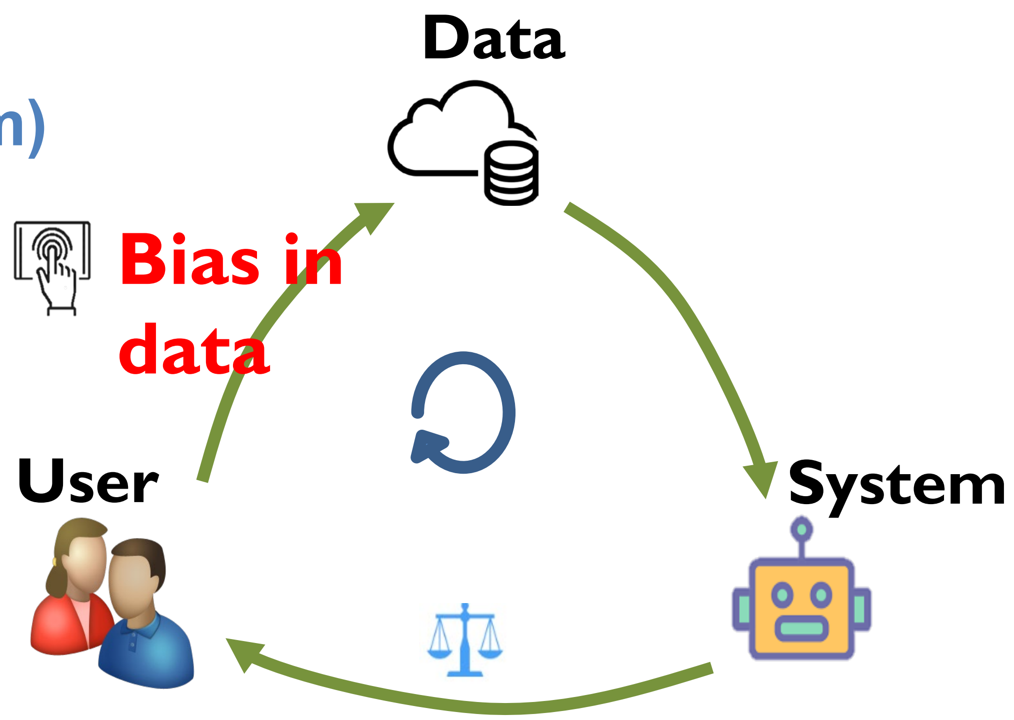
- RS captures user preference and provides **personalized information filtering**
- **However, bias widely exists in RS**

Bias in Recommendation

Bias in Data (collecting)

- Data is observational rather than experimental (i.e., missing-not-at-random)
- Labels do not faithfully reflect user preference
- Affected by many factors:
  - Exposure mechanism
  - Public opinions
  - Display position

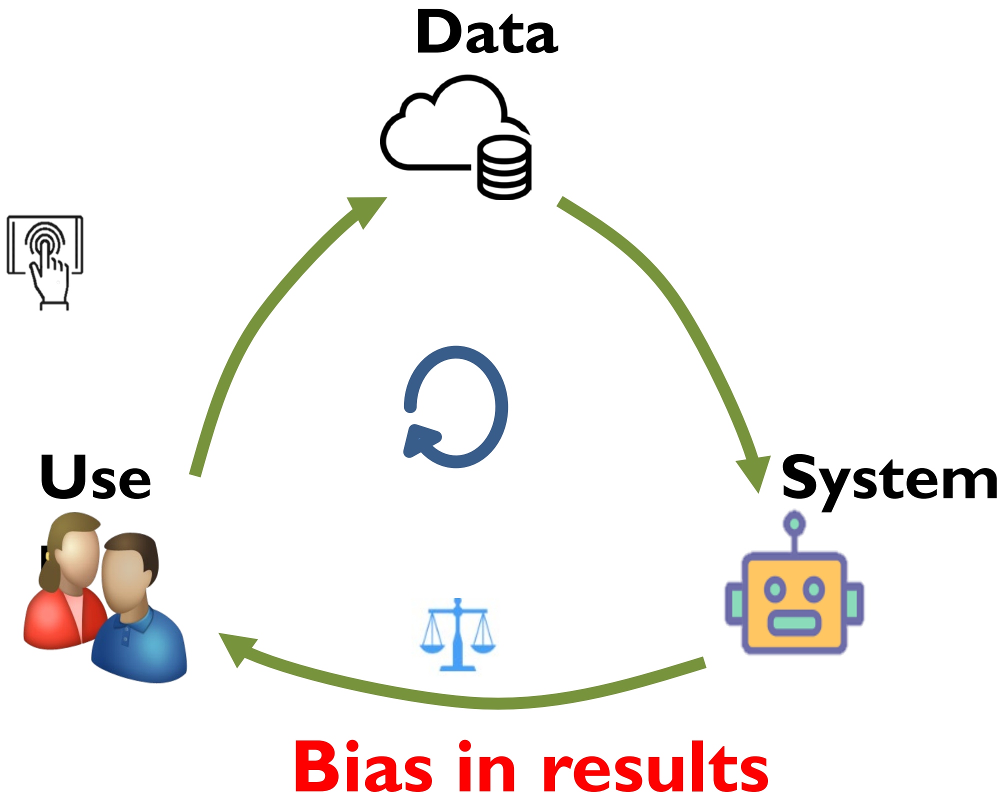
  …..
- The collected data deviates from user true preference



Bias in Recommendation

☐ Bias in Results (training & serving)

- Models do not only inherit the bias in data but also amplify/generate bias
- Models naturally bias towards dominated groups
- E.g. Popularity bias, mainstream bias



Bias is Evil

☐ Economic perspective

- Decreasing recommendation accuracy
- Hurting user experience and satisfaction
- Causing the losses of users

- **Example: The clickbait problem.** Items with interesting title (but boring content) may get more exposure opportunity. But users do not like them!

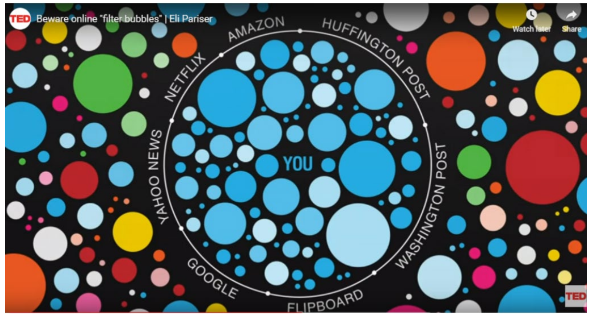
Chen et al. Bias and Debias in Recommender System: A Survey and Future Directions. TOIS' 22.

Bias is Evil

☐ Social perspective

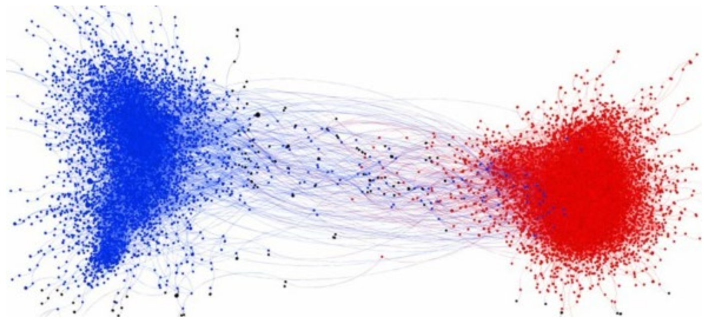
Unfairness

E.g., job recommendation [1]:
- software developer → man
- registered nurse → woman



Filter Bubble

E.g., Tittytainment
- Entertainment videos/games
- Addictive



Polarization

E.g., political polarization [1]:
- Democrats vs Republicans

Debiasing is vitally important!

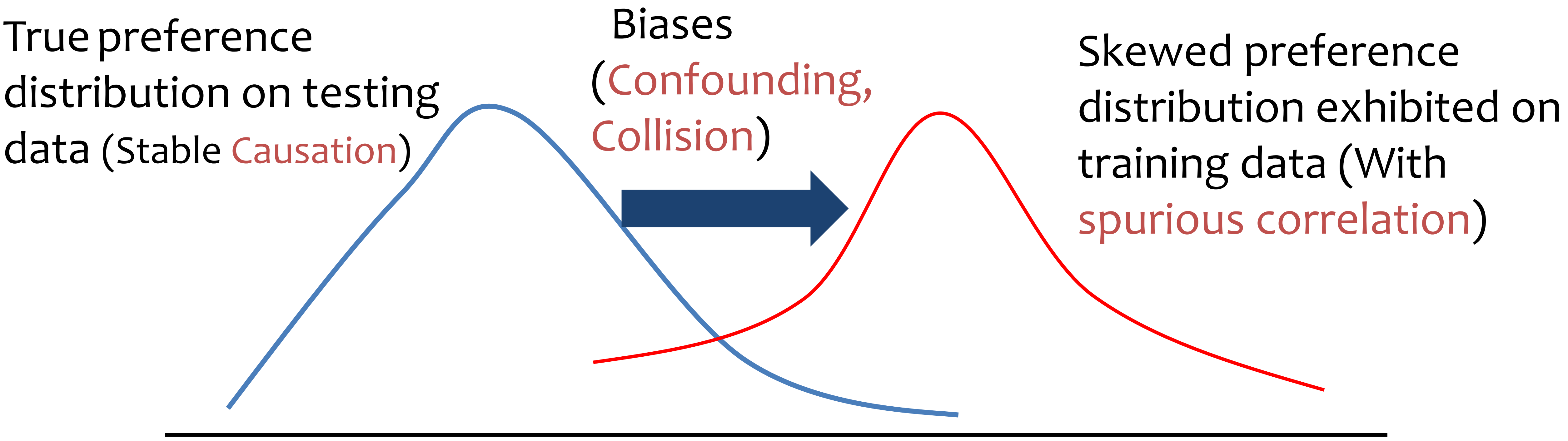
Milano et.al. "Recommender systems and their ethical challenges." Ai & Society 35.4 (2020): 957-967.

❖ Outline

1. Bias in Recommendation
2. **Our Recent Debiasing Work**
   - Survey on Recommendation Bias
   - MACR （借鉴）
   - TIDE（改进）
   - UnKD （新问题）
   - Adap-τ (分析)

Survey on Recommendation Bias(对问题总结和归纳)

- 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.

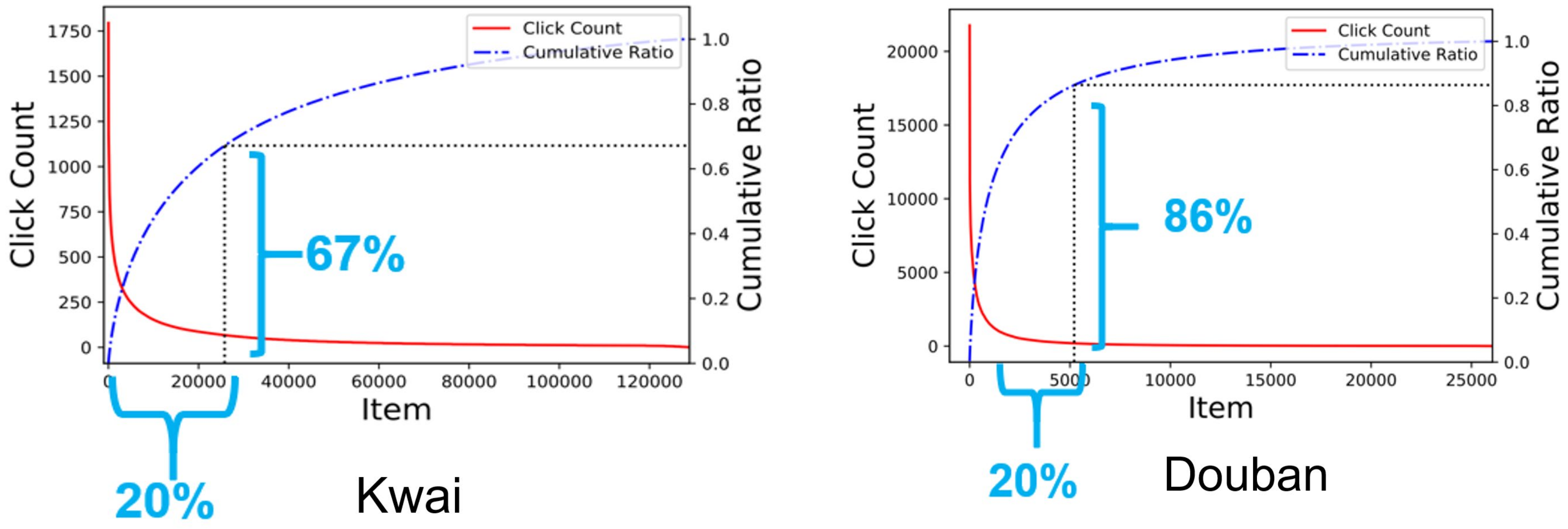
Chen et al. TOIS'22. Bias and Debias in Recommender System: A Survey and Future Directions

• Debiasing Strategies Overview

1. For Selection Bias
   - Data imputation (Relabeling)
   - Propensity score (Reweighting)
   - Doubly robust model (Reweighting+Relabeling)
   - Generative model (Generative modeling)
2. For Conformity Bias
   - Disentangling (Generative modeling)
3. For Exposure Bias
   - Downweighting (Reweighting+Relabeling)
   - Exposure-based generative model (Generative modeling)
4. For Position Bias
   - Click model (Generative modeling)
   - Propensity score (Reweighting)
   - Trust-aware model (Reweighting+Relabeling)

Popularity Bias

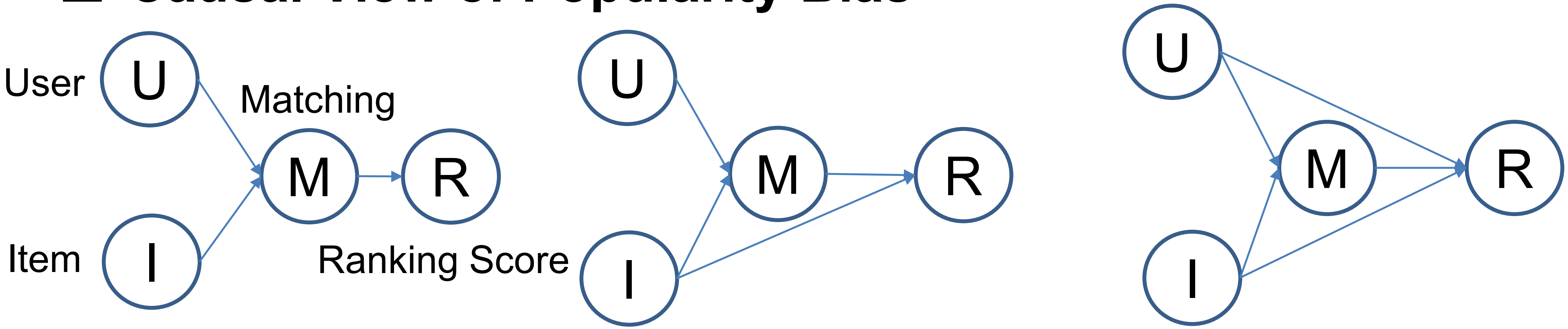
- **Popularity bias in recommender system**
  - **Popularity bias**: Popular items are recommended even more frequently than they would warrant
    - Long-tail distribution, Matthew effect
    - Harmful for personality and accuracy



Long-tail distribution

MACR: Model-Agnostic Counterfactual Reasoning (借鉴)

☐ 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' sensitivity to popularity.**

☐ Solution Idea:

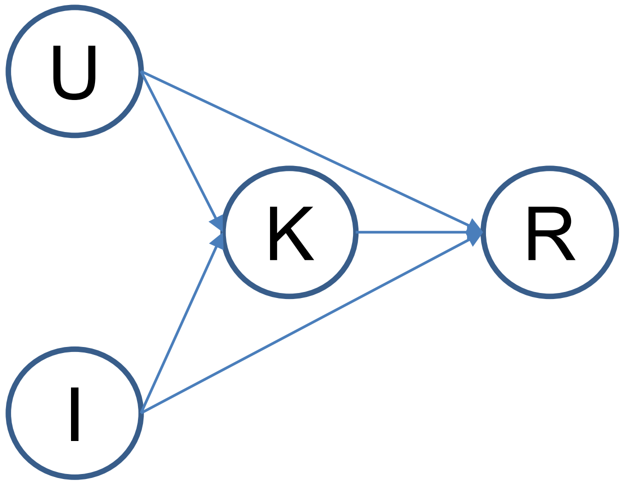
- Train a recommender based on the causal graph via a **multi-task learning**
- Perform **counterfactual inference** to eliminate **popularity bias**

Question to answer: what would the prediction be if there were only popularity bias?

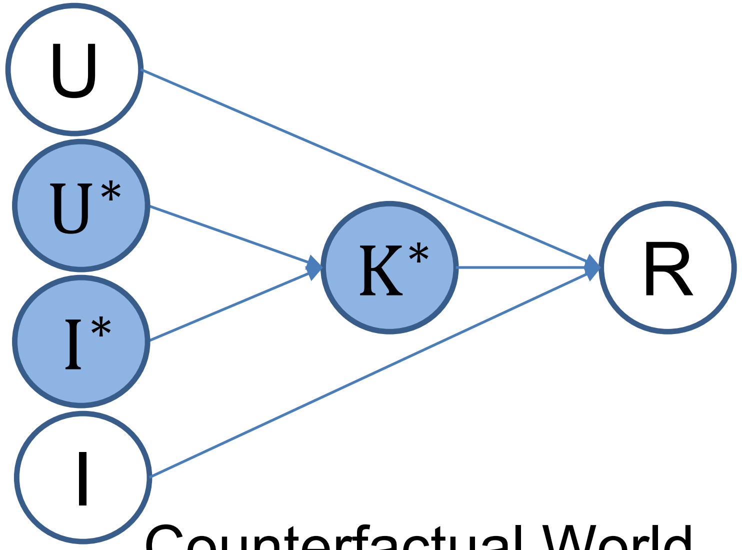
Wei et al. **Model-agnostic Counterfactual Reasoning for Eliminating Popularity bias in Recommender System**. KDD' 21.

MACR: Model-Agnostic Counterfactual Reasoning

- **Counterfactual Inference to Remove Bias**



Factual World
(original prediction)



Counterfactual World
(block matching to capture bias)

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

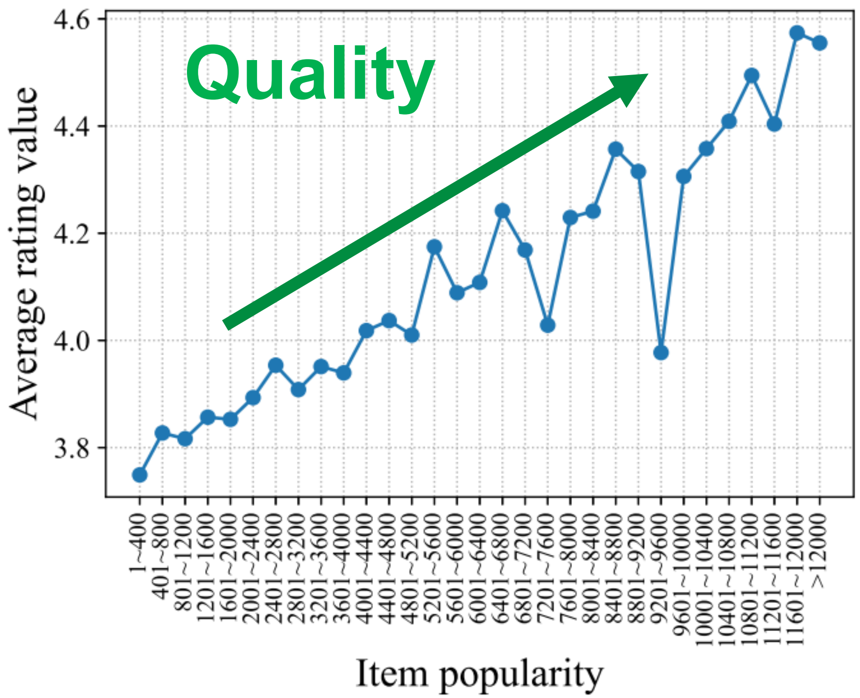
Total Indirect Effect | Total Effect | Natural Direct | Factual world | Counterfactual world

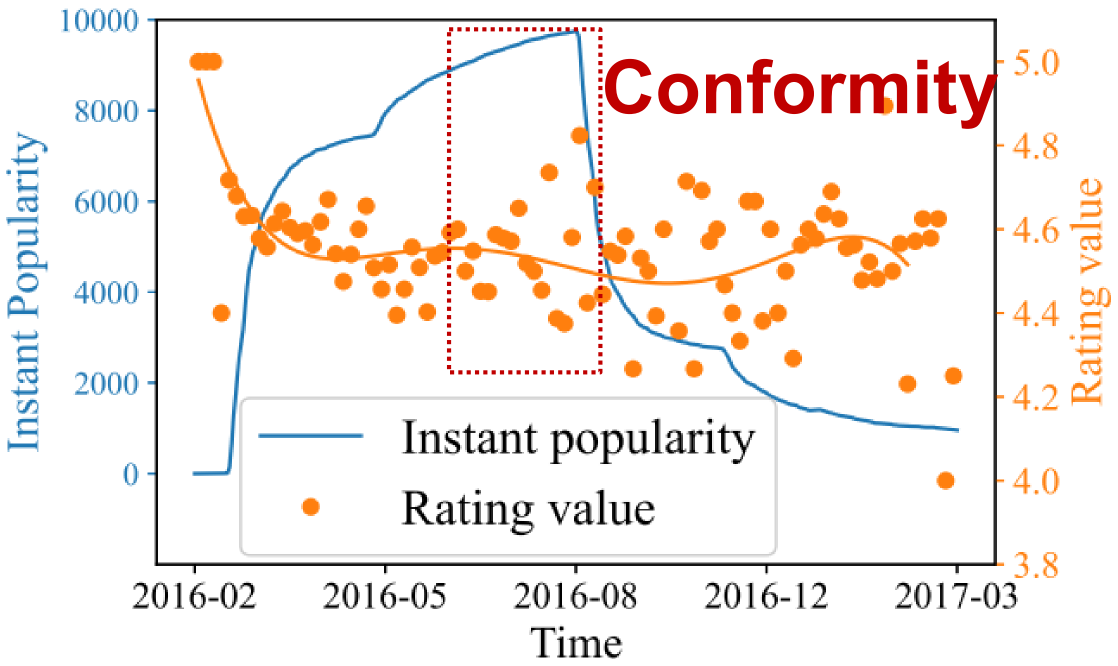
Inference with TIE $= \hat{y}_k \times \sigma(\hat{y}_i) \times \sigma(\hat{y}_u) — c \times \sigma(\hat{y}_i) \times \sigma(\hat{y}_u)$

Wei et al. **Model-agnostic Counterfactual Reasoning for Eliminating Popularity bias in Recommender System**. KDD' 21.

TIDE: Disentangling Benign and Harmful Bias （改进）

- **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

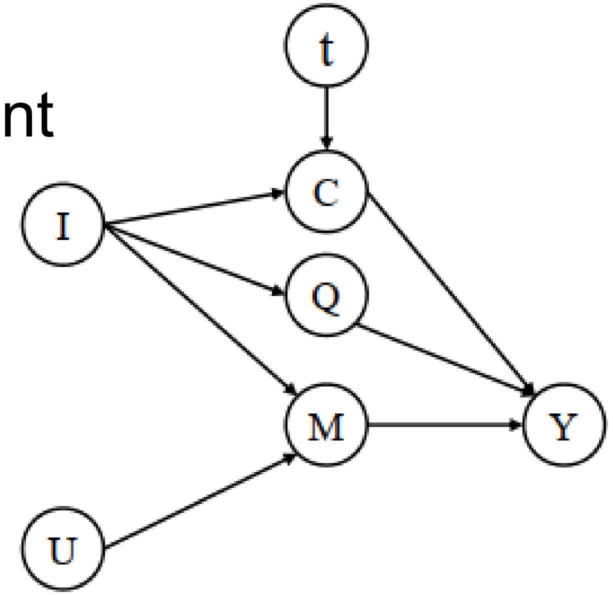




Zhao et al. **Popularity Bias is not Always Evil: Disentangling Benign and Harmful Bias for Recommendation**. TKDE' 22.

TIDE: Disentangling Benign and Harmful Bias

- Time-aware **D**is**E**ntangled framework(**TIDE**)
  - Main challenge: Lack of explicit signal for disentanglement



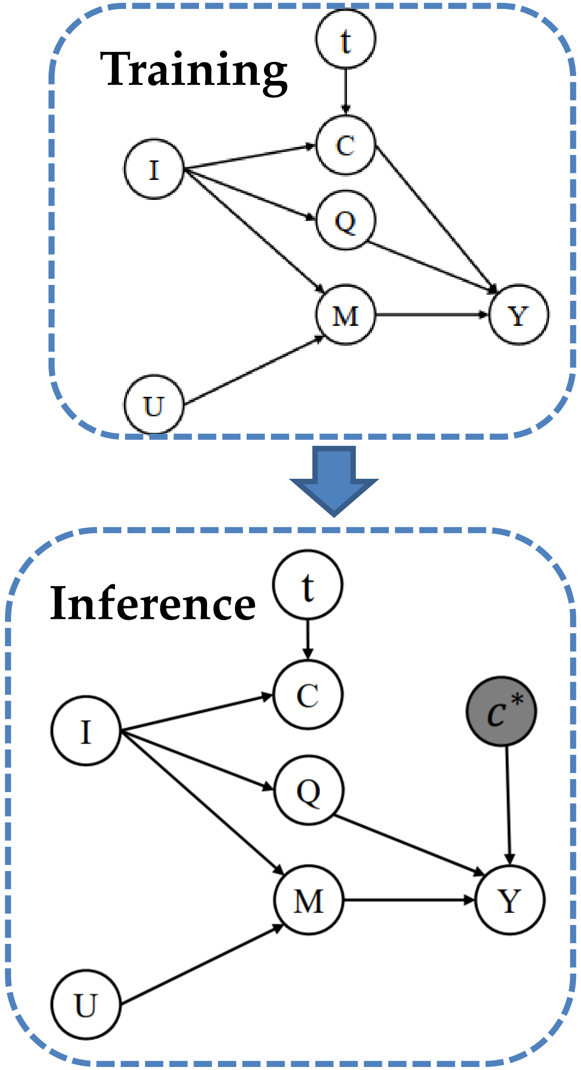
(a) Causal graph of our TIDE.

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

- **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**, since recent interactions should have stronger influence.
- **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.

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

TIDE: Disentangling Benign and Harmful 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:

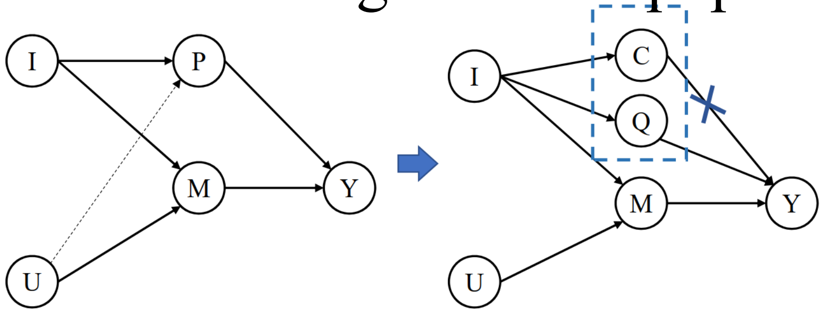
- Intervention: set c as reference vector c^* (e.g., zero) during inference to **remove** the **improper effect from C to Y**.

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

☐ Comparison with PD

- TIDE further conduct disentanglement of popularity bias

PD

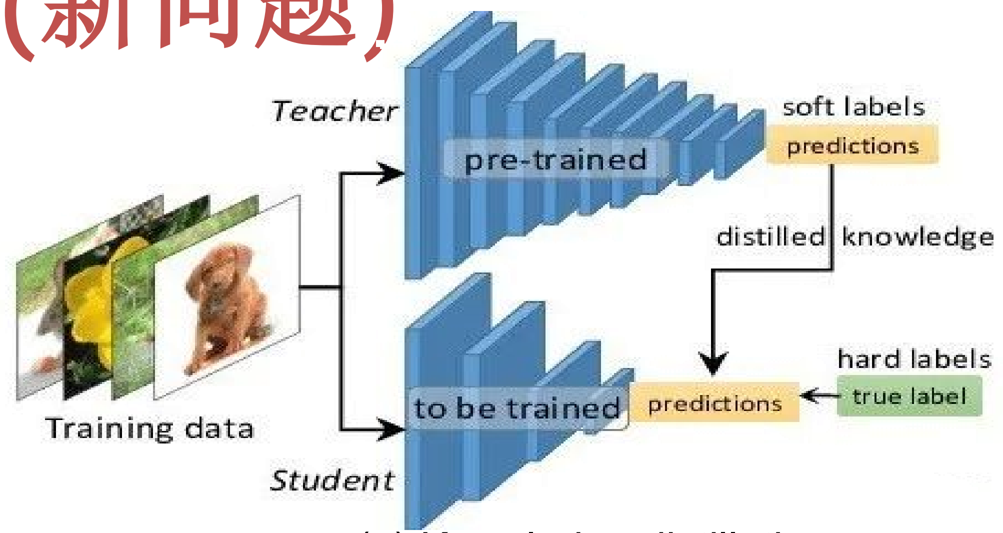


TIDE

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

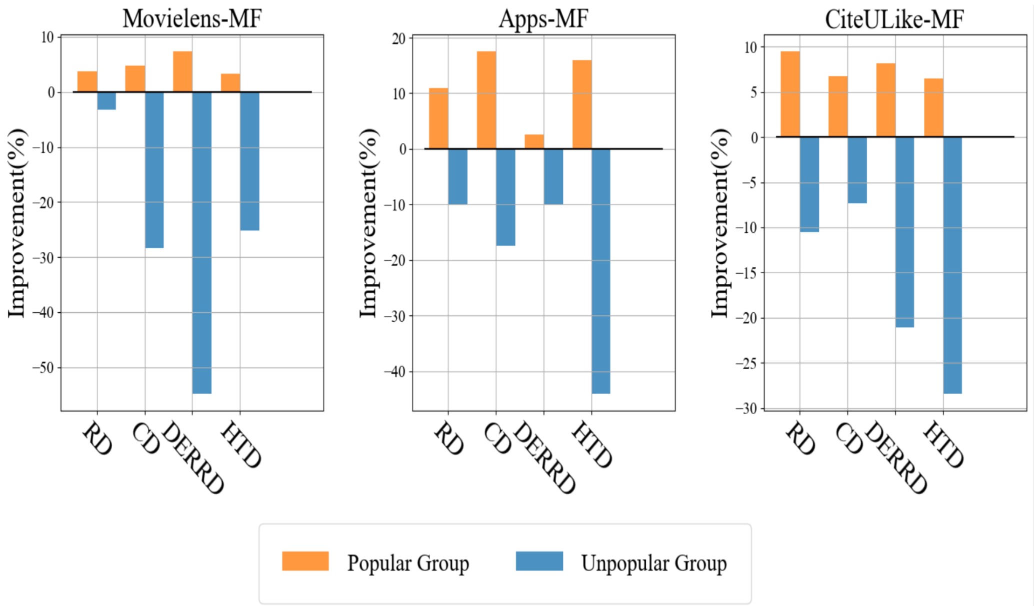
UnKD: Unbiased Distillation (新问题)

- **Knowledge distillation:** Transfer knowledge from large models to small models.



(a) Knowledge distillation

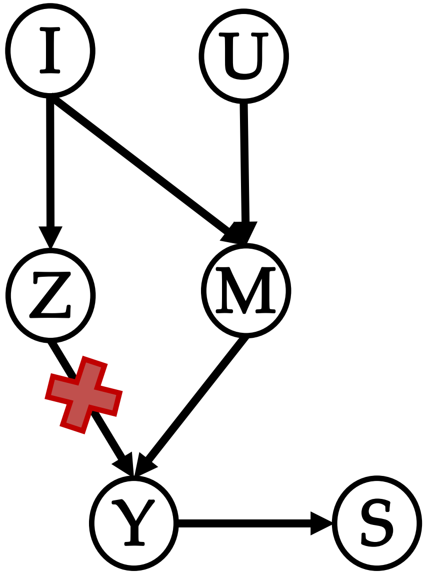
- **Existing knowledge distillation:**
  - Performance is mainly improved in the popular group
  - The performance of the unpopular group was severely degraded.



(b) The improvement ratio of different distillation methods.

Gang Chen et al. Unbiased Knowledge Distillation for Recommendation. WSDM' 23.

UnKD: Unbiased Distillation



- **Conditional total effect :** For a particular user u,

  $TE_i = Y_{i|u} - Y_{i^*|u}$; i^* is the benchmark situation

- **Adverse causal effect** : $PEZ_i = Y_{i^*,Z_i|u} - Y_{i^*|u}$

- **Eliminating bias :** $PEM_i = TE_i - PEZ_i = Y_{i|u} - Y_{i^*,Z_i|u}$

- **For any two items *i* and *j* with highly similar popularity**, we have $Z_i \approx Z_j$, so the equation $Y_{i^*,Z_i|u} = Y_{i^*,Z_j|u}$ almost holds. Then, we have:

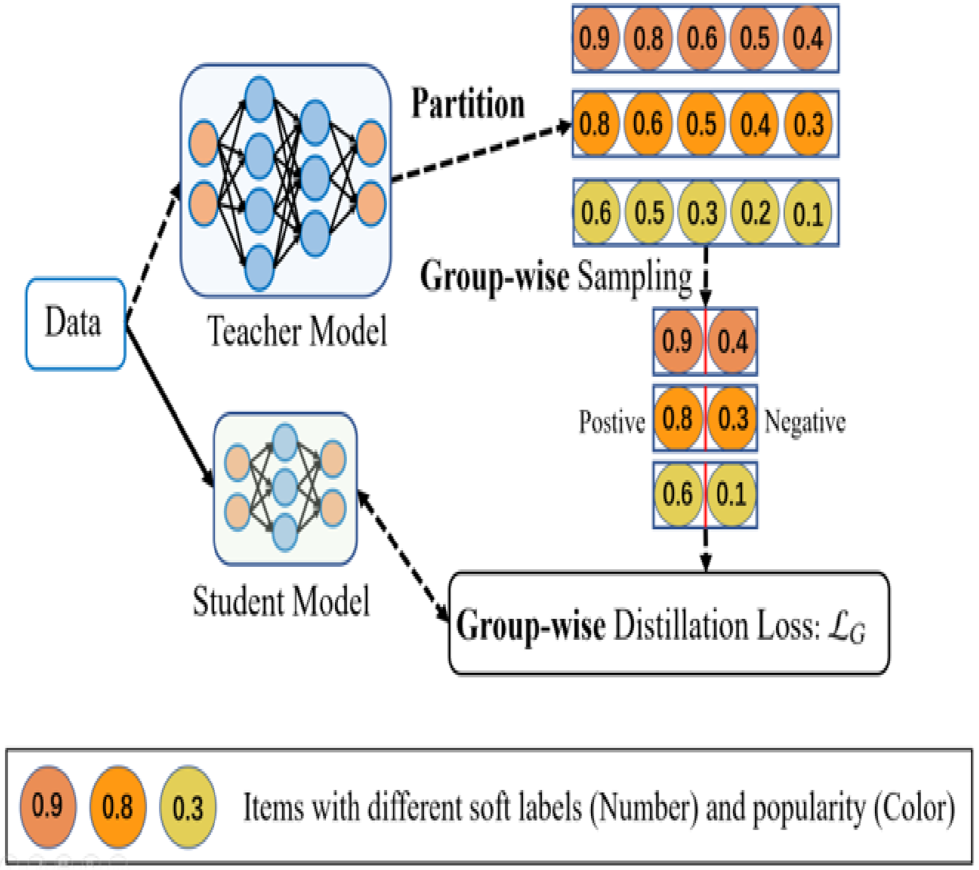
$$Y_{i|u} > Y_{j|u} \Leftrightarrow Y_{i|u} - Y_{i^*,Z_i|u} > Y_{j|u} - Y_{i^*,Z_j|u} \Leftrightarrow PEM_i > PEM_j$$

Gang Chen et al. Unbiased Knowledge Distillation for Recommendation. WSDM' 23.

UnKD: Unbiased Distillation

- **UnKD consists of the following three steps:**
  - **Group partition.** Items are divided into K groups according to their popularity.
  - **Group-wise Sampling**. Sample a set of $\mathcal{S}_{ug}$ positive and negative item pairs for each group $g(i^+, i^-)$.
  - **Group-wise Learning**. The student model is trained with a group distillation loss：

$$\mathcal{L}_G = -\sum_{u} \frac{1}{|\mathcal{U}|} \sum_{g \in \mathcal{G}} \sum_{(i^+, i^- \in \mathcal{S}_{ug})} log\ \sigma\left(\boldsymbol{e}_u^T \boldsymbol{e}_{i^+} - \boldsymbol{e}_u^T \boldsymbol{e}_{i^-}\right)$$



(a) UnKD Structure

Gang Chen et al. Unbiased Knowledge Distillation for Recommendation. WSDM' 23.

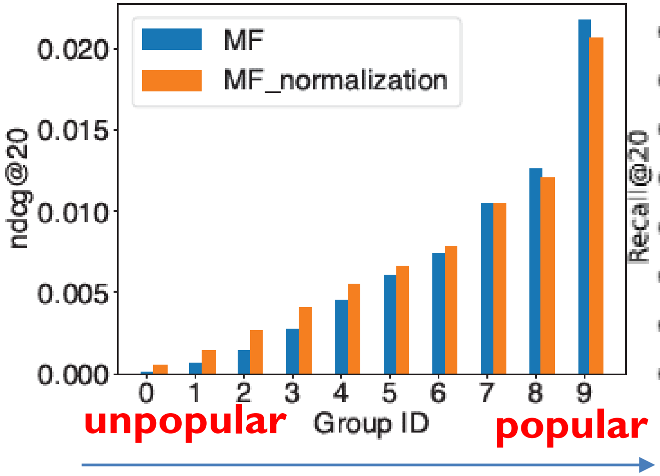
Adap-τ: Embedding Normalization(发现+分析问题)

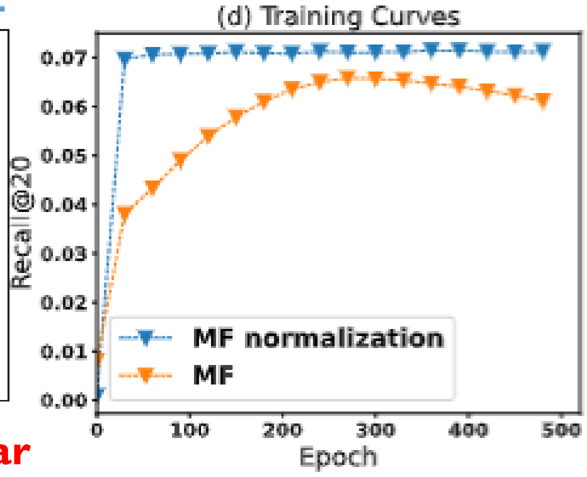
inner product: $f_{u,i} = e_u^T \cdot e_i$

normalization: $f_{u,i} = \frac{e_u^T \cdot e_i}{\|e_u\| \cdot \|e_i\|} \cdot \frac{1}{\tau}$

□ Harm of Un-normalized embedding:

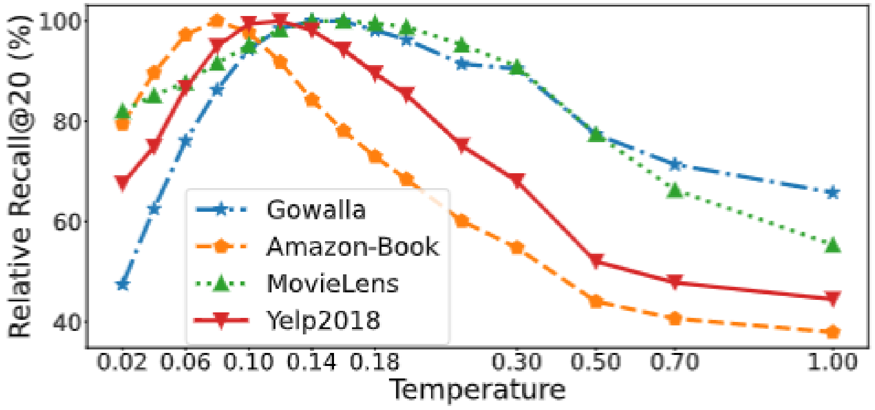
- **Aggravates popularity bias**
- **Hurts training convergence**





□ Challenge of embedding normalization:

- highly sensitive to the hyper-parameter temperature τ.
- Finding proper τ is difficult.



Wu et al. Adpatively Modulating Embedding Magnitude for Recommendation. Submit to WWW'23.

Adap-τ: Embedding Normalization

☐ Understanding roles of temperature:

- R1: Avoiding from gradient vanishment

  => temperature should adapt to the data and model

- R2: Balancing contributions from hard instances and easy instances

  => temperature could be fined-grained (personalized) for flexible adjustment

- **We propose adaptive and fine-grained temperature:**
  - Adaptive: maximizing the cumulated gradient magnitude:

$$\boxed{\tau_0 \approx \frac{\mu_+ - \mu}{\log(\frac{nm}{2|D|})}}$$

μ_+: average score of positive instances
μ: average score of all instances
n: the number of users
m: the number of items
D: the number of positive instances

  - Fine-grained: monitoring the loss for each user:

$$\boxed{\tau_u^* = \tau_0 \cdot \exp(\mathbb{W}(\max(-\frac{1}{e}, \frac{L(u) - m_u}{2\beta})))}$$

$L(u)$: the loss for the user u
m_u: mean of $L(u)$
W: lambert-W function

Wu et al. Adpatively Modulating Embedding Magnitude for Recommendation. Submit to WWW'23.

Conclusion and Future Work

- **Idea? 合适的方法+有用的问题**
    - 合适的方法：借鉴，改进
    - 有用的问题：寻找新的问题（可解决的）
    - 分析型文章：研究方法和问题的性质
- **Future direction: 以动态图表征为例**
    - 合适的方法：自监督学习、Transformer、改进效率
    - 有用的问题：鲁棒性、可解释性、隐私保护、动态图应用
    - 分析型：表征的几何结构?

THANK YOU!