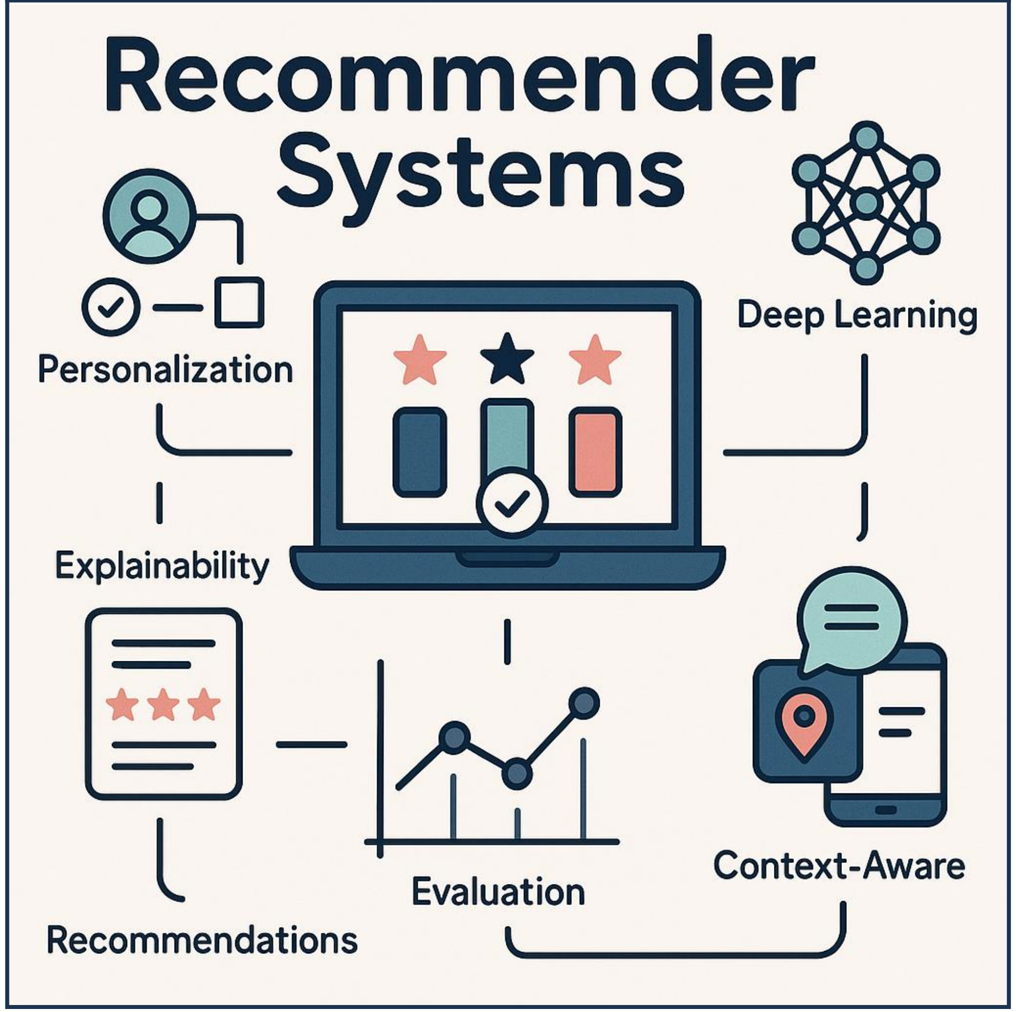
Generate an image that represents the current state of research in recommender systems



--- by ChatGPT

NANYANG TECHNOLOGICAL UNIVERSITY

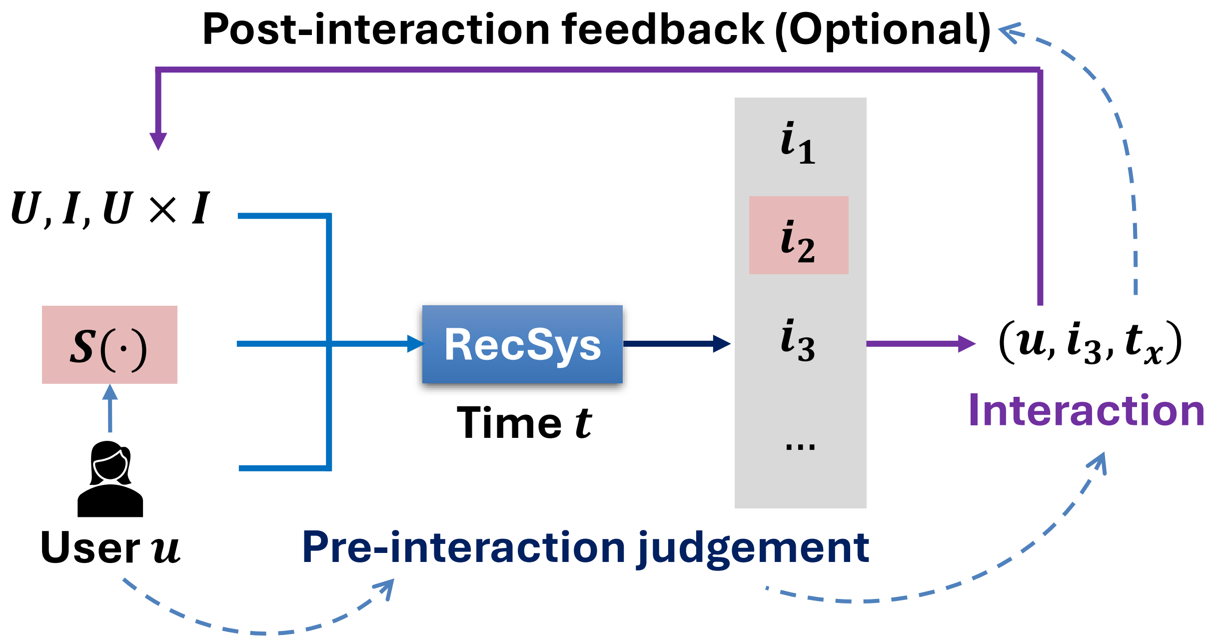
From Recommendations to Interactions:

Putting Users Back in the Loop

Dr. Aixin Sun 孙爱欣

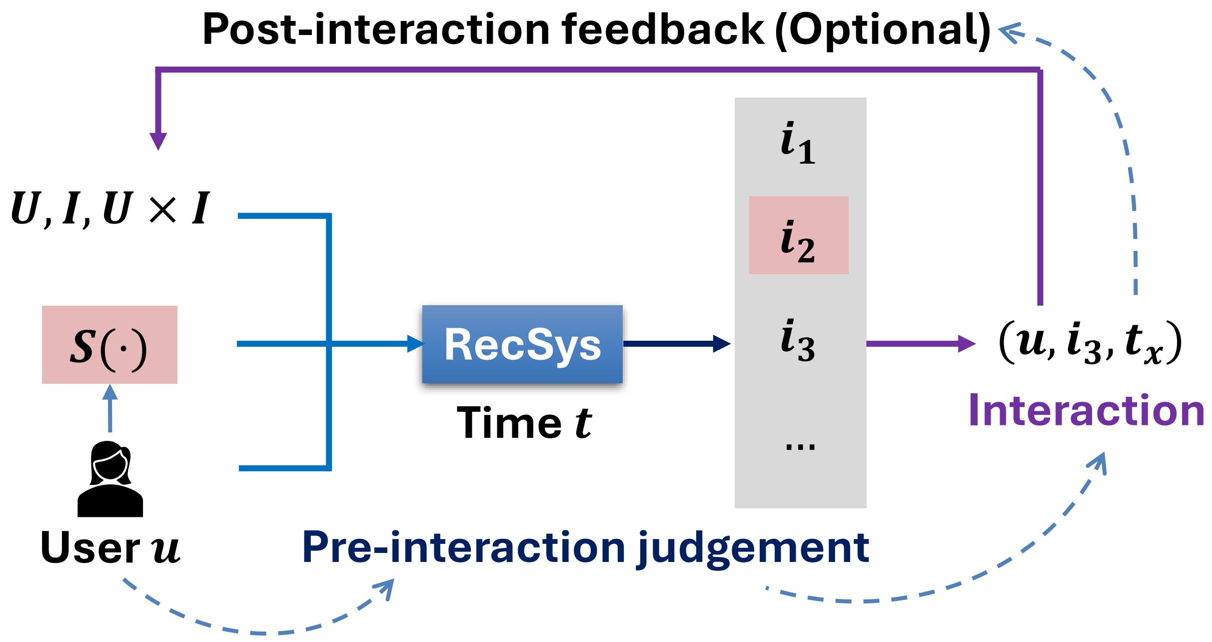
NTU Singapore

Users U, Items I, and User-Item Interactions $U \times I$



- RecSys models learn from **user-item interactions**, and/or ratings/reviews
  - To learn user preference
  - The goal is to reduce **user cost** (but **what are those costs**?)
- **Assumption:** if a user interacts with an item, it reflects some form of preference for that item.

User-Item Interactions $U \times I$: The life cycle



- **Pre-interaction judgment:** When user is presented with a list of recommended items
- **Interaction**: When user interacts with a chosen item
- **Post-interaction**: Rating/review

Hotel booking/staying as an example

- **Pre-interaction judgment**
  - A user is presented with a list of 10 hotels on a booking website
  - A decision influenced by factors such as images, branding, location, price, or other readily available attributes.
  - Further efforts: Reads the hotel description, checks room facilities, and reviews user feedback before deciding to book.
  - **Cost**: the user's **time and effort** in gathering relevant information

- **Interaction**
  - Stays at the hotel and gains firsthand experience.
- **Post-interaction**
  - Provides feedback in the form of a rating and review
  - **Cost**: additional effort of providing review/rating

- Primary reasons for choosing this hotel
- Stay experience and genuine preference
- Gap between the expectations at booking and the true experiences

Image: Google Image Search

Complexity of Pre-Interaction Judgment

- **Informed** vs **Uninformed** Decision
  - Whether user has the knowledge to accurately judge an item before interacting with it?
  - Familiar items like books, movies, or other products the user has prior experience
  - If a user has never used a robot before, many of the terms in the product description may be unfamiliar to them, even after reading user reviews.

- **Uninformed decisions may not necessarily indicate user preferences**.
  - It is challenging to determine whether a user's decision was based on prior knowledge or made without a full understanding of the item.

Image: https://robotsguide.com/robots/optimus

Items in One vs. Multiple Types/Categories

- RecSys for **one type of items**, e.g., books, movies, news, or music.
  - There are **common characteristics**, such as genre, director, or artist, that users can rely on for pre-interaction judgment before actually interacting with a recommended item

- RecSys for **items of multiple types**/categories, e.g., e-commerce
  - Users apply different **criteria and expectations** when making judgments for different types of items, for informed decisions.
  - Shall user preference be based on item types/categories?
  - **User preference** vs general **associative patterns**

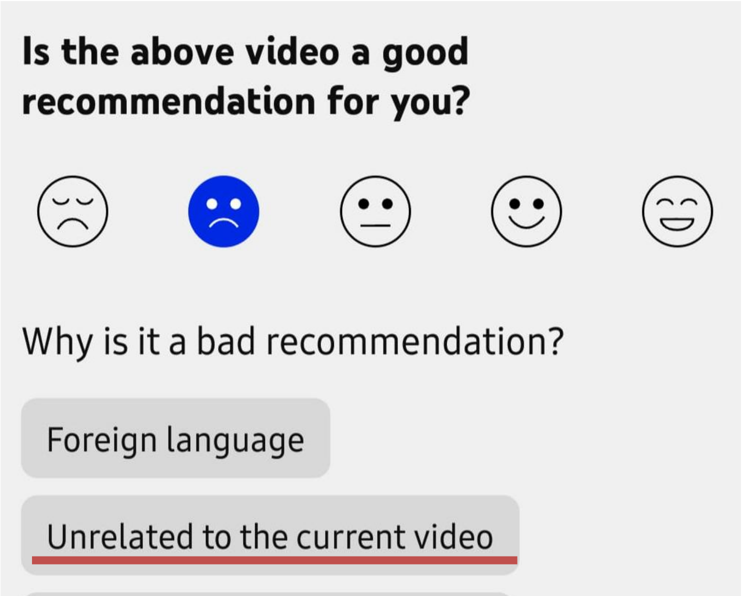
Recognition of User-Item Interaction

- User-item interactions are recorded as (u, i, t_x) in most datasets. The interaction process can be complicated, e.g., E-commerce
  - Add to cart, make payment, receive delivery, and return?
  - Should a return be considered a valid interaction for learning user preference?

- Absence of Pre-Interaction Judgment
  - Music streaming and short-video viewing
    - Users often do not actively select each item
  - User feedback?
    - User engagement signals, such as skipping, fast forwarding, or continuing to watch/listen
    - User tolerance?

Recognition of User-Item Interaction

- User-item interactions are recorded as (u, i, t_x) in most datasets
- **Unobservable** Interaction
  - Job recommendation
  - CV + Skills vs Job opening → applied job? Received an offer? Accepted the offer?
- **Interdependency** across recommendations



Recommendation vs User Cost

- The ultimate goal of a recommender system is
  - to **reduce user effort** in finding products or services of interest
  - to **enhance their enjoyment** of recommendations
  - to **build trust** in the system

- **Different costs** at **various stages** of the interaction process
  - Pre-interaction judgment stage
  - Interaction stage
  - Post-interaction stage

  - Shall all forms of cost being considered in model/loss function design?
  - Are the costs the same for different recommendation scenarios/applications?

Task, Solution, and Evaluation

- Is the RecSys research task **well defined**?

- $(U, U \times I, I) \rightarrow R^u$

TABLE 1
A Fragment of a Rating Matrix for a Movie Recommender System

	K-PAX	Life of Brian	Memento	Notorious
Alice	4	3	2	4
Bob	∅	4	5	5
Cindy	2	2	4	∅
David	3	∅	5	2

Items I

User/Item	i_1	i_2	i_3	i_n
u_1	✓		✓	
u_2		✓		
u_3	✓	?	?	✓
u_m		✓		✓

Interactions $U \times I$

734 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 6, JUNE 2005

Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions

Gediminas Adomavicius, *Member, IEEE*, and Alexander Tuzhilin, *Member, IEEE*

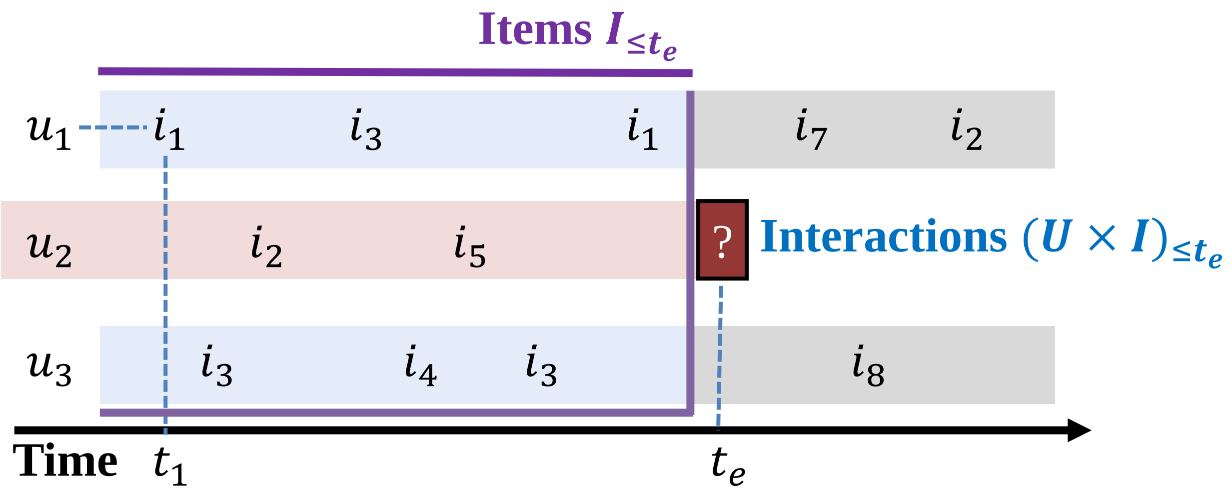
Abstract—This paper presents an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches. This paper also describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multcriteria ratings, and a provision of more flexible and less intrusive types of recommendations.

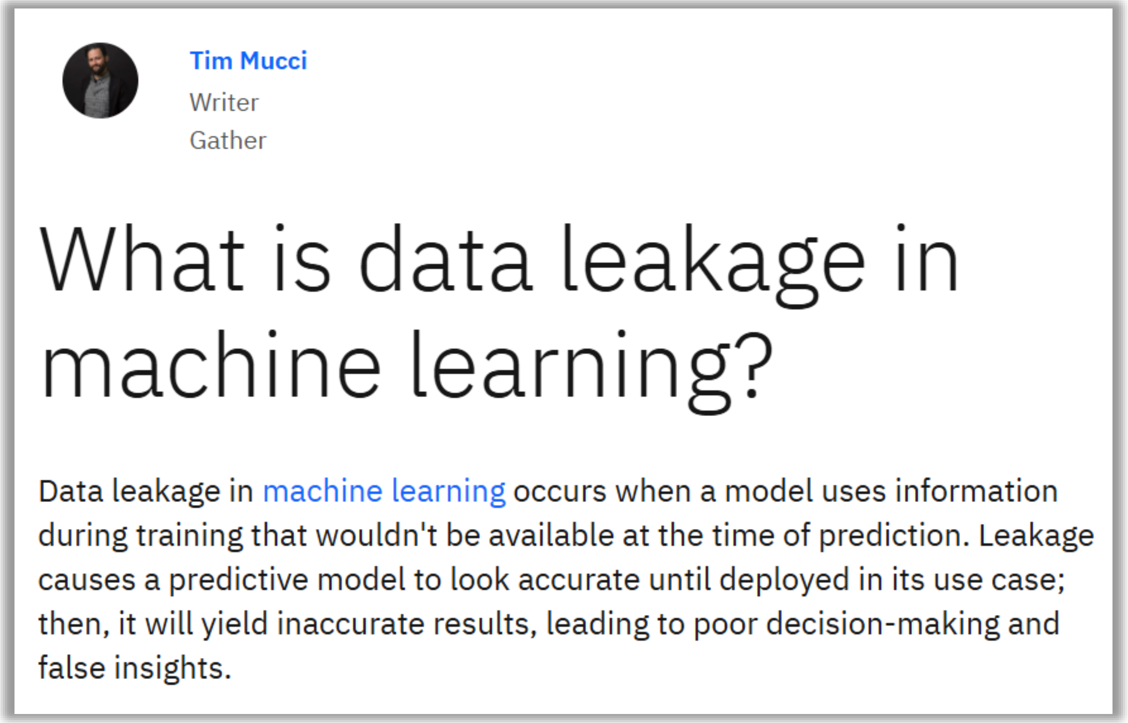
Index Terms—Recommender systems, collaborative filtering, rating estimation methods, extensions to recommender systems.

Task, Solution, and Evaluation

- Is the RecSys research task **well defined**?

- $(U, t, (U \times I)_{\leq t}, I_t) \rightarrow R_t^u$




Tim Mucci
Writer
Gather

What is data leakage in machine learning?

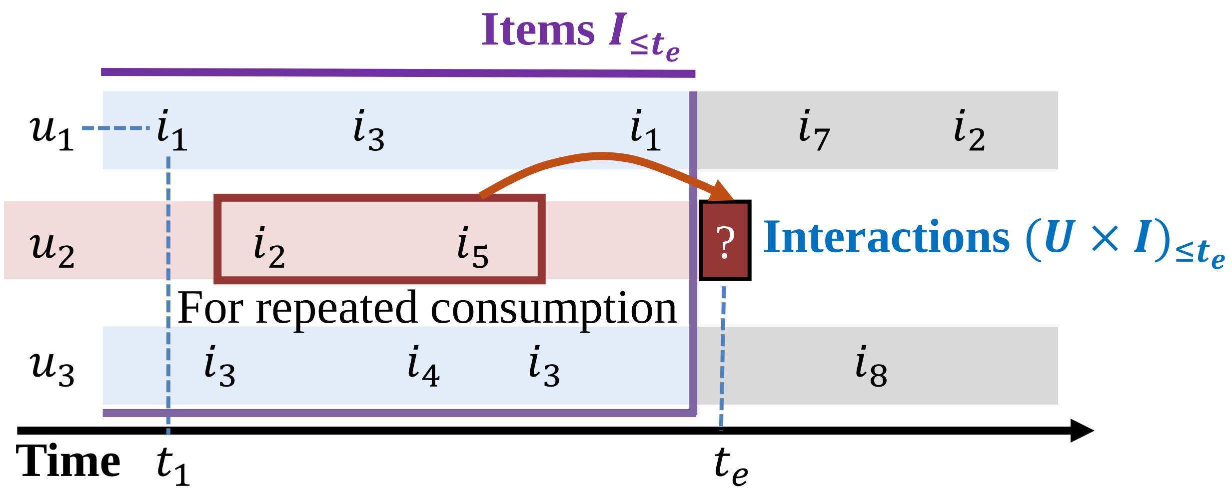
Data leakage in machine learning occurs when a model uses information during training that wouldn't be available at the time of prediction. Leakage causes a predictive model to look accurate until deployed in its use case; then, it will yield inaccurate results, leading to poor decision-making and false insights.

https://www.ibm.com/think/topics/data-leakage-machine-learning

Task, Solution, and Evaluation

Repeated consumption?

- Is the RecSys research task **well defined?**



Many other item selection criteria by users: e.g., geo-distance

Impact: offline evaluation

When considering all these factors

Table 1: Example recommendation tasks across different domains. The characteristics are based on our best understanding, which may not be applicable to any specific platform. The "price" attribute is ignored from pre-interaction judgment.

Domain	Item Type	Session	Candidate Item Selection	Pre-Interaction Judgment	User-Item Interaction
Music streaming	Single	Yes	Repeat, Exploration	No, except the first song	Progress ratio, Skip
Short video	Single	Yes	No, or non-repeat	No, except the first video	Progress ratio, Watching speed, Skip
Online shopping	Multi	Might	Repeat, Exploration	Description, Image, Review	Payment, Return
News	Single	Yes	Region, Genre	Title, Source	Click, View
Job	Single	No	Location	Description	Unobservable, Post-interaction inference
Course	Single	No	Meeting pre-requisite	Description, Instructor, Institute	Progress, completion
Food delivery	Single	No	Repeat, Exploration, Delivery distance/time	Image, Description, Review	Payment
POI	Multi	Might	Region, Repeat, Exploration	Description, Image, Travel cost, Review	Unobservable, Post-interaction inference

The reality in academic research

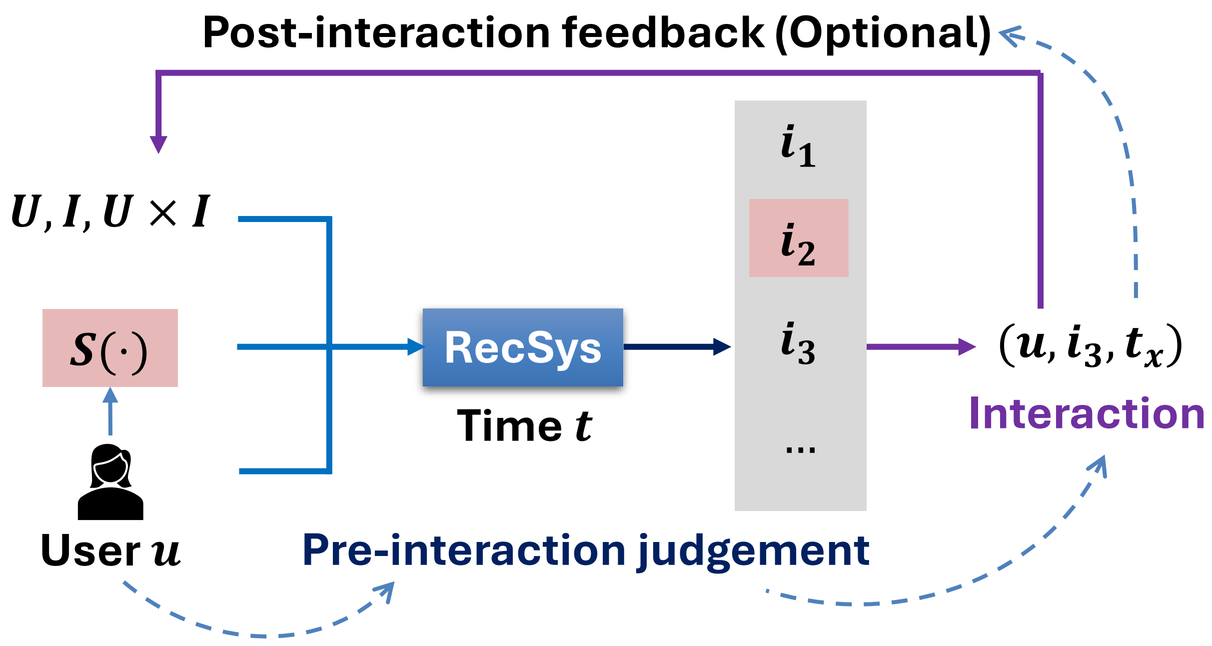
https://arxiv.org/abs/2410.22972

Table 1: Datasets, pre-processing and splitting strategies adopted in 55 research papers published between 2020 and 2024. Only datasets and strategies appearing in at least 3 research papers are shown.

55 papers

			Datasets																Pre-processing							Splitting						
Paper	Venue	Year	Yelp	Amazon Books	MovieLens 1M	Amazon Beauty	Last.fm	Gowalla	Amazon Toys and Games	Tmall	Amazon Games	Yelp2018	Alibaba-iFashion	Amazon Clothing	Amazon Sports and Outdoors	Epinions	MovieLens 20M	*Others*	*None*	Binarization	*k*-Core Iterative	*k*-Core User	*k*-Core Item	Session length	*Others*	Random Hold-Out	Temporal Leave-1-out Hold-Out	Fixed	Random Leave-1-out Hold-Out	Temporal Fixed Handcrafted	*Others*	*NA*
Hao et al. [31]	TOIS	2023			✓			✓										✓	✓							✓						
Shuai et al. [62]	SIGIR	2022	✓						✓					✓				✓				✓				✓						
Jiang et al. [40]	KDD	2023	✓				✓											✓	✓		✓					✓						
Yu et al. [93]	SIGIR	2022		✓								✓						✓	✓	✓						✓						
Xia et al. [82]	WWW	2023	✓	✓				✓											✓							✓						
Hansen et al. [30]	RecSys	2020																✓	✓											✓		
Kowald et al. [44]	Inf. Sci.	2021																✓							✓						✓	
Yu et al. [90]	TKDE	2022										✓	✓					✓	✓							✓						
			17	14	10	9	9	8	7	7	4	4	3	3	3	3	3	40	25	15	13	9	4	4	7	21	13	5	4	3	10	1

Putting Users Back in the Loop



https://personal.ntu.edu.sg/axsun/

Cornell University

Computer Science > Information Retrieval

arXiv:2503.21188 (cs)

[Submitted on 27 Mar 2025 (v1), last revised 15 Apr 2025 (this version, v2)]

Are We Solving a Well-Defined Problem? A Task-Centric Perspective on Recommendation Tasks

Aixin Sun

View PDF

HTML (experimental)

Recommender systems (RecSys) leverage user interaction history to predict and suggest relevant items, shaping user experiences across various domains. While many studies adopt a general problem definition, i.e., to recommend preferred items to users based on past interactions, such abstraction often lacks the domain-specific nuances necessary for practical deployment. However, models are frequently evaluated using datasets from online recommender platforms, which inherently reflect these specificities. In this paper, we analyze RecSys task formulations, emphasizing key components such as input-output structures, temporal dynamics, and candidate item selection. All these factors directly impact offline evaluation. We further examine the complexities of user-item interactions, including decision-making costs, multi-step engagements, and unobservable interactions, which may influence model design and loss functions. Additionally, we explore the balance between task specificity and model generalizability, highlighting how well-defined task formulations serve as the foundation for robust evaluation and effective solution development. By clarifying task definitions and their implications, this work provides a structured perspective on RecSys research. The goal is to help researchers better navigate the field.

https://arxiv.org/abs/2503.21188