




# Chapter 6: Recommender Systems


崔金华

电子邮箱: [jhcui@hust.edu.cn](mailto:jhcui@hust.edu.cn)

个人主页: <https://csjhcui.github.io/>


# Example: Recommender System



**Jinhua Cui (崔金华)** 


Huazhong University of Science and Technology  
在 hust.edu.cn 的电子邮件经过验证 - 首页

Flash Memory SSD Storage System Approximate Storage

 已关注


<input type="checkbox"/> 标题	引用次数	年份
<input type="checkbox"/> <a href="#">A Space-Efficient Fair Cache Scheme Based on Machine Learning for NVMe SSDs</a> W Liu, J Cui, T Li, J Liu, LT Yang IEEE Transactions on Parallel & Distributed Systems, 1-18		2022
<input type="checkbox"/> <a href="#">Improving 3D NAND SSD Read Performance by Parallelizing Read-Retry</a> J Cui, Z Zeng, J Huang, W Yuan, LT Yang IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems	1	2022
<input type="checkbox"/> <a href="#">IMRSim: A Disk Simulator for Interlaced Magnetic Recording Technology</a> Z Zeng, X Chen, LT Yang, J Cui arXiv preprint arXiv:2206.14368		2022
<input type="checkbox"/> <a href="#">Exploiting Uncorrectable Data Reuse for Performance Improvement of Flash Memory</a> J Cui, C Liu, J Liu, J Huang, LT Yang IEEE Transactions on Computer-Aided Design of Integrated Circuits and ...	2	2021
<input type="checkbox"/> <a href="#">ADS: Leveraging Approximate Data for Efficient Data Sanitization in SSDs</a> J Cui, W Liu, J Huang, LT Yang IEEE Transactions on Computer-Aided Design of Integrated Circuits and ...	3	2021
<input type="checkbox"/> <a href="#">MLCache: A space-efficient cache scheme based on reuse distance and machine learning for NVMe SSDs</a> W Liu, J Cui, J Liu, LT Yang 2020 IEEE/ACM International Conference On Computer Aided Design (ICCAD), 1-9	7	2020
<input type="checkbox"/> <a href="#">Exploiting disturbance-aware read redirection for performance improvement in 3D flash memory</a>	3	2020

Google 学术搜索




☒ 不限语言 ☐ 中文网页 ☐ 简体中文网页

推荐的文章





[Improving the Endurance of Next Generation SSD's using WOM-v codes](#)  
S Jaffer, K Mahdavian, B Schroeder  
ACM Transactions on Storage - 7 天前






[Accelerating Time Series Analysis via Processing using Non-Volatile Memories](#)  
I Fernandez, A Manglik, C Giannoula, R Quisilant, NM Ghiasi, J Gó...  
arXiv preprint arXiv:2211.04369 - 13 天前





[Dynamic Error Recovery Flow Prediction Based on Reusable Machine Learning for Low Latency NAND Flash Memory under Process Variation](#)  
M Hwang, J Jee, J Kang, H Park, S Lee, J Kim  
IEEE Access - 14 天前





[Multi-resource fair allocation for consolidated flash-based caching systems](#)  
W Choi, B Urgaonkar, MT Kandemir, G Kesidis

## Customer X

- Published papers
- Saved papers in personal library

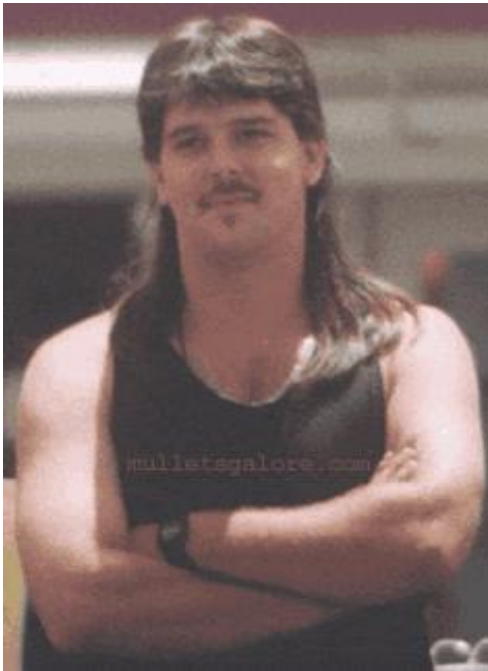
## Customer X

- Recommend related papers

# Another Example: Recommender System

## □ Customer X

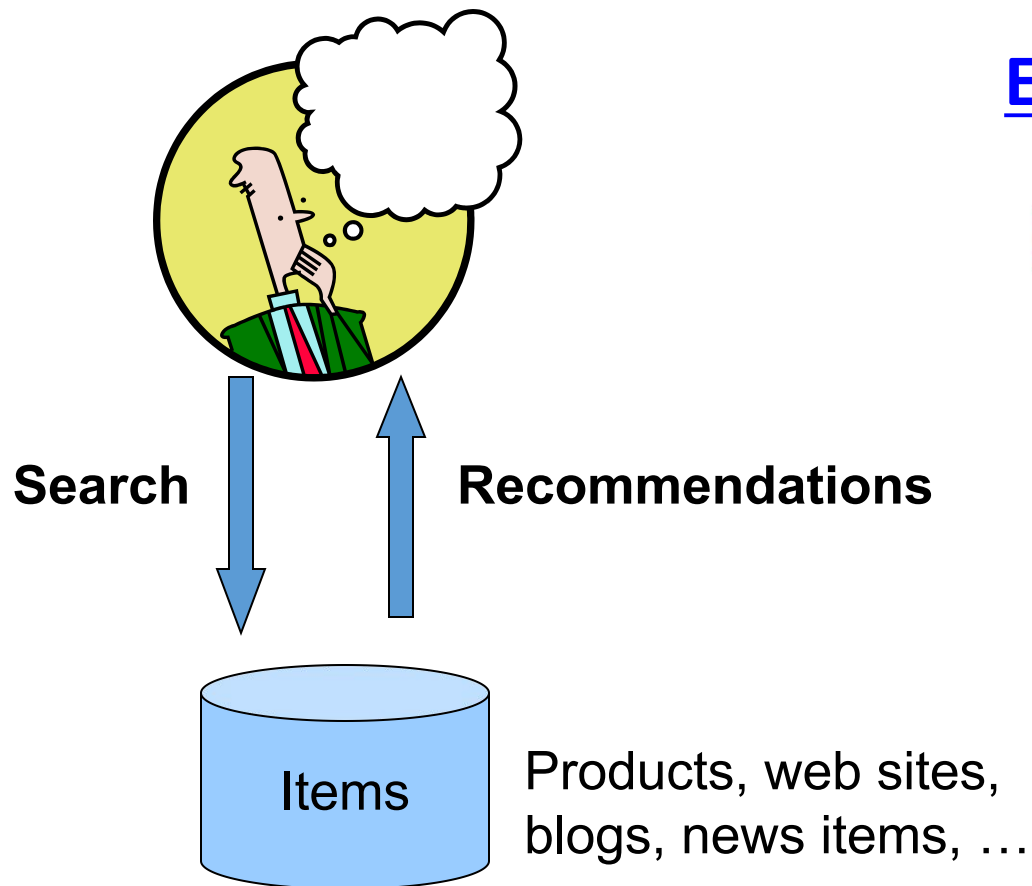
- Buys Metallica(金属乐队) CD
- Buys Megadeth(麦格德斯) CD



## □ Customer Y

- Does search on Metallica
- Recommender system suggests Megadeth from data collected about customer X

# Recommendations



## Examples:

Baidu 百度

抖音

知乎  
有问题 上知乎

bilibili

芒果tv

新浪微博

淘宝网  
Taobao.com

豆瓣 douban

爱奇艺

腾讯视频

YOUKU

...

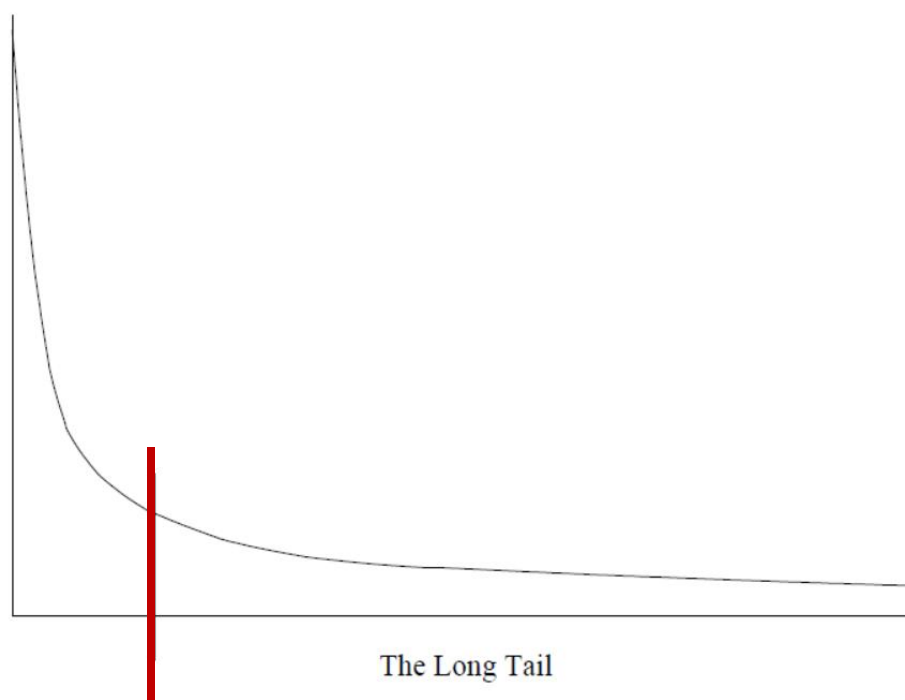


# From Scarcity to Abundance

- ❑ **Shelf space is a scarce commodity for traditional retailers**
  - Also: TV networks, movie theaters,...
- ❑ **Web enables near-zero-cost dissemination of information about products**
  - From scarcity to abundance
  - **Long tail phenomenon (长尾现象)**

# The Long Tail

□ The distinction between the physical and on-line worlds has been called the **long tail phenomenon (长尾现象)**.



- 纵坐标代表流行度 (the number of times an item is chosen).
- 所有项按照流行度在横坐标上排序.
- 实体机构只列出红色竖线左边的流行项; 在线机构能提供包括流行项和尾部项在内的全范围的项.

# From Scarcity to Abundance

- **Shelf space is a scarce commodity for traditional retailers**
  - Also: TV networks, movie theaters,...
- **Web enables near-zero-cost dissemination of information about products**
  - From scarcity to abundance
  - **Long tail phenomenon (长尾现象)**
- **More choice necessitates better filters**
  - Recommendation engines
  - How **Into Thin Air** 《**巅峰**》 made **Touching the Void** 《**攀越冰峰**》 a bestseller: <http://www.wired.com/wired/archive/12.10/tail.html>

# Types of Recommendations

## ❑ Editorial and hand curated

- List of favorites
- Lists of “essential” items

## ❑ Simple aggregates

- Top 10, Most Popular, Recent Uploads

## ❑ Tailored to individual users

- Amazon, Netflix, ...



□  $X$  = set of **Customers**

□  $S$  = set of **Items**

□ **Utility matrix** (效用矩阵)  $u. X \times S \rightarrow R$

➤  $R$  = set of ratings

➤  $R$  is a totally ordered set

➤ e.g., **0-5** stars, real number in **[0,1]**

# Utility Matrix

	Avatar (阿凡达)	LOTR (指环王)	Matrix (黑客帝国)	Pirates (加勒比海盗)
Alice	1		2	
Bob		5		3
Carol	2		1	
David				4

## □(1) Gathering “known” ratings for matrix

- How to collect the data in the utility matrix

## □(2) Extrapolate unknown ratings from the known ones

- Mainly interested in high unknown ratings
  - We are not interested in knowing what you don't like but what you like

## □(3) Evaluating extrapolation methods

- How to measure success/performance of recommendation methods

# (1) Gathering Ratings

## □ Explicit

- Ask people to rate items
- Doesn't work well in practice – people can't be bothered

## □ Implicit

- Learn ratings from user actions
  - E.g., purchase implies high rating
- What about low ratings?

## ■ Hybrid: both explicit and implicit

## (2) Extrapolating Utilities

### □ **Key problem:** Utility matrix $U$ is **sparse**

- Most people have not rated most items
- **Cold start:**
  - New items have no ratings
  - New users have no history

### □ **Approaches to recommender systems:**

- **1)** Content-based
- **2)** Collaborative
- **3)** Latent factor based
- .....