

# A Brief Introduction to Recommender Systems

Presenter: Liang Hu

# Outline

- Recommendation age
- A case study of recommender systems
- The 3C principle to build recommender systems
- Classic recommender systems
- Evaluation metrics for recommender systems

# Content

- Recommendation age
- A case study of recommender systems
- The 3C principle to build recommender systems
- Classic recommender systems
- Evaluation metrics for recommender systems

# Information age

## Connecting world with searches

Google search results for "recommender systems":

About 4,160,000 results (0.55 seconds)

en.wikipedia.org › wiki › Recommender\_system

**Recommender system - Wikipedia**

A recommender system, or a recommendation system (sometimes replacing 'system' with a synonym such as platform or engine), is a subclass of information filtering system that seeks to predict the "rating" or "preference" a user would give to an item. They are primarily used in commercial applications.

Category:Recommender ... · Cold start (recommender ... · Collaborative filtering

You've visited this page many times. Last visit: 8/2/18

towardsdatascience.com › recommender-systems-in-prac...

**Recommender Systems in Practice - Towards Data Science**

Feb 13, 2019 - Companies like Amazon, Netflix, LinkedIn, and Pandora leverage recommender systems to help users discover new and relevant items ...

People also ask

- Which algorithms are used in recommender systems?
- Where are recommender systems used?
- How does a recommender system work?
- Why do we need recommender systems?

Feedback

towardsdatascience.com › introduction-to-recommende...

**Introduction to recommender systems - Towards Data Science**

Jun 2, 2019 - In a very general way, recommender systems are algorithms aimed at suggesting relevant items to users (items being movies to watch, text to ...)

Amazon search results for "recommender systems":

1-48 of 505 results for "recommender systems"

Department

- Books
- Computers & Technology
- Databases & Big Data
- Artificial Intelligence Expert Systems
- Data Mining
- Artificial Intelligence & Semantics
- See more
- Kindle Store
- Computers & Technology
- Computer Databases
- AI & Semantics
- Expert Systems
- Data Mining
- See more
- See All 3 Departments

Avg. Customer Review

- ★★★★★ & Up
- ★★★★☆ & Up
- ★★★☆☆ & Up
- ★★☆☆☆ & Up

Computer User Books

- Advanced & Power Users
- Beginners & Seniors

Book Format

- Hardcover
- Paperback
- Kindle Edition

Amazon Global Store

- Amazon Global Store

Condition

- New

Feedback

Programming Pragmatics

Gear Up Your Programming Skills

Engineering: A Compiler

Agile Systems Engineering

The Art and Science of Analyzing Software Data

Recommender Systems

Practical Recommender Systems

Building Recommender Systems with Machine Learning and AI: Help people discover new products and content with deep learning, neural networks, and more!

Recommender Systems: An Introduction

Hands-On Recommendation Systems with Python: Start building powerful and personalized, recommendation...

# Time changes

## Connecting world with clicks



This screenshot shows a mobile application interface. At the top, there is a search bar with the placeholder "搜索感兴趣的内容" and a red "热搜" button. Below the search bar, there are several news cards:

- A card about "习近平在陕西平利考察脱贫攻坚情况" with a timestamp of 01:41.
- A card about "总书记告诫的“大教训”是什么?" with a timestamp of 14:15.
- A card about "美国十多个州暴发骚乱：示威者武装比特种部队还精良，不要封城要自由" with a timestamp of 11:59.
- A card about "个人条款基本达成一致，曼联将以破英超转会纪录的方式击败切尔西" with a timestamp of 14:15.
- A card about "低至45元！屁股不闷热，比空调还凉快！" with a timestamp of 14:15.
- A card about "小游戏 象棋的常见套路行不通了，快来救残局！" with a timestamp of 14:15.

At the bottom of the screen, there is a navigation bar with icons for "首页" (Home), "视频" (Videos), "菜单" (Menu), "小视频" (Short Videos), and "我的" (My Profile).

This screenshot shows a mobile application interface. At the top, there is a search bar with the placeholder "b站大会员". Below the search bar, there are several sections:

- A "关注" (Follow) section.
- A "推荐" (Recommendation) section.
- Category links: 食品 (Food), 水果 (Fruit), 男装 (Men's Clothing), 手机 (Mobile Phone), 百货 (General Merchandise).
- A "拼小圈" (Pinduoduo Circle) section.
- A "秒杀" (Flash Sale) section.
- A "断码清仓" (Clearance) section.
- A "上海发货" (Shanghai Shipment) section.
- A "免费领水果" (Free Fruit) section.
- A "9块9特卖" (Special Offer) section.
- A "充值中心" (Recharge Center) section.
- A "医药馆" (Pharmacy) section.
- A "签到" (Check-in) section.
- A "多多赚大钱" (DuoDuo Zuan Daqian) section.
- A "免费领商品" (Free Product) section.
- A "多多买菜" (DuoDuo Mai Cai) section.
- Product categories: 新鲜果蔬 (Fresh Fruits and Vegetables), 肉禽蛋品 (Meat, Poultry, Eggs), 乳品冲饮 (Dairy Products), 粮油调味 (Grains, Oils, Condiments).
- A "百亿补贴" (Billions of Yuan Subsidy) section.
- Product details: Dove chocolates (¥4779), a smartphone (¥5290), a massage gun (¥4779), and a small item (¥118).

At the bottom of the screen, there is a navigation bar with icons for "首页" (Home), "多多视频" (DuoDuo Videos), "分类" (Categories), "聊天" (Chat), and "个人中心" (Personal Center).

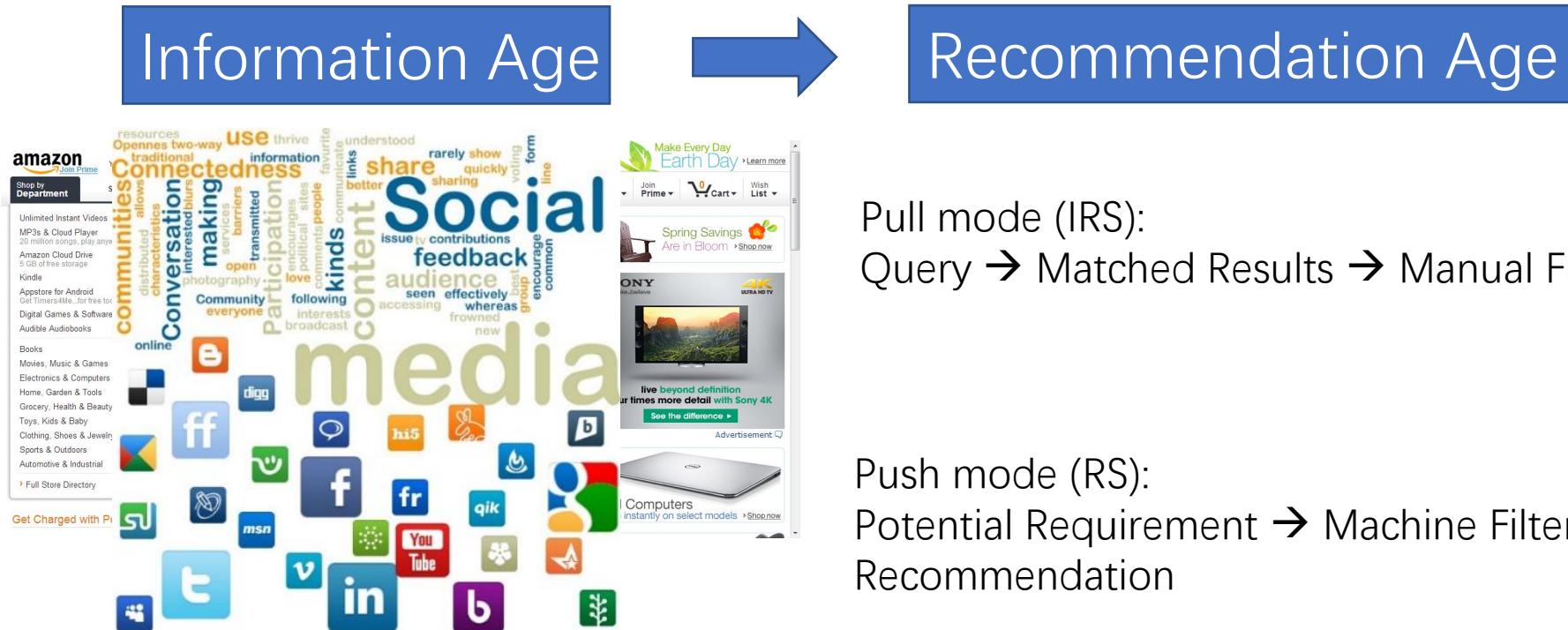
This screenshot shows a mobile application interface for a video platform. At the top, there is a search bar with the placeholder "b站大会员". Below the search bar, there are several video thumbnails and their details:

- [Thumbnail] 【实况足球】进入腾讯后第一件事送掉网易号, ...  
观看数: 1.9万, 评论数: 44, 时长: 2:40
- [Thumbnail] 大家换老婆的速度, 在下叹为观止!  
观看数: 1563, 评论数: 3, 时长: 2:29
- [Thumbnail] 【国足·点赞飙升】  
观看数: 27.9万, 评论数: 682, 时长: 2:51
- [Thumbnail] 鞭炮虾? 北京牛? 国外最火的中餐厅都在卖啥?  
观看数: 10.2万, 评论数: 1582, 时长: 6:19
- [Thumbnail] 现在迎面走来的是我的救命恩人方队——玄冥女...  
观看数: 1.9万, 评论数: 1582, 时长: 6:19
- [Thumbnail] 美食圈 · 中餐厅  
观看数: 1.9万, 评论数: 1582, 时长: 6:19
- [Thumbnail] 17斤胖喵吓坏医生! 秤都要撑不住了  
观看数: 203.6万, 评论数: 93.2万
- [Thumbnail] 那个...真的要亲一下吗?  
观看数: 1.9万, 评论数: 1582, 时长: 6:19
- [Thumbnail] 纪录片 · 猫咪体检记  
观看数: 1.9万, 评论数: 1582, 时长: 6:19
- [Thumbnail] 游戏 · 忍耐下就好  
观看数: 1.9万, 评论数: 1582, 时长: 6:19
- [Thumbnail] 史上最强世界杯  
观看数: 1.9万, 评论数: 1582, 时长: 6:19
- [Thumbnail] 王校长嫌我胖跑了~  
观看数: 1.9万, 评论数: 1582, 时长: 6:19

At the bottom of the screen, there is a navigation bar with icons for "首页" (Home), "频道" (Channels), "动态" (Activities), and "会员购" (Membership Purchase).

# From information age to recommendation age

- Recommender systems (push information) are the evolution of information retrieval systems (pull information).



Pull mode (IRS):  
Query → Matched Results → Manual Filtering

Push mode (RS):  
Potential Requirement → Machine Filtering → Recommendation

# Personalization



# Recommender systems have occupied our life

## What to eat

### Food & Drink

Sort by **Relevance**

[View On Map](#)



**Up to 67% off Room Hire with Drinks for Four**  
Dynasty Karaoke  
Up to ten party-goers can add more zing to their night out with a private room hire featuring a glass of house wine, beer or soft drink each

Haymarket • 2.2 km  
[View Deal](#)



**Up to 60% off Japanese BBQ Special for Two**  
Taiso Wagyu Japanese BBQ  
Zetland • 5 km  
90+ bought  
[\\$98 From \\$58](#)



**Up to 43% off French Fine Dining with Cocktails**  
The Little Snail  
The Little Snail • 1.7 km  
[\\$154 From \\$88](#)



**Up to 57% off Darts + Beer or Mix Drinks for 2**  
The Century Bar  
Sydney • 1.9 km  
[\\$44 From \\$19](#)



**Up to 57% off Festive Buffet + Delivery**  
CNI Catering  
[\\$299.50 From \\$149](#)

## Which to dress

Sort by **Relevance**



**Men's Compression Shorts**  
20+ bought [From \\$15](#)



**Skechers Shoes for Women and Men**  
70+ bought [From \\$59](#)



**Customised Build-A-Brick Cap**  
30+ bought [From \\$19](#)



**Women's Knitted Ugg Boots**  
90+ bought [\\$219 From \\$65](#)



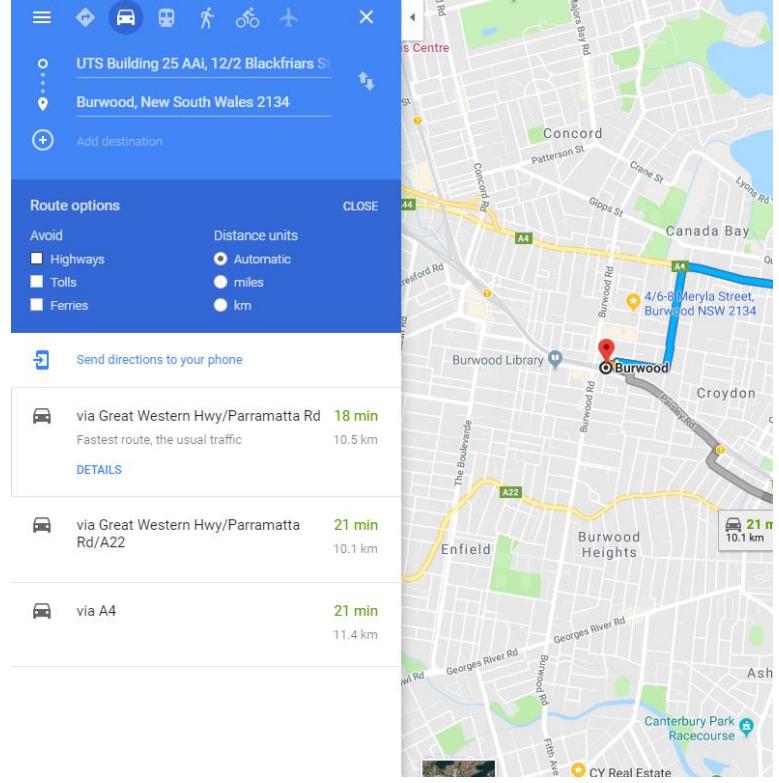
**Ralph Lauren Polo Shirt**  
250+ bought [\\$120 From \\$59](#)



**Two-Piece Men's Thermal Wear Set**  
20+ bought [From \\$25](#)

## Where to go

Sort by **Relevance**



**Route options**

Avoid  
 Highways  
 Tolls  
 Ferries

Distance units  
 Automatic  
 miles  
 km

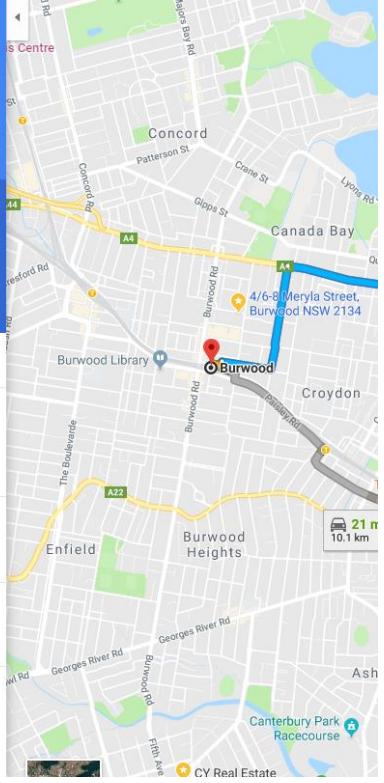
[CLOSE](#)

[Send directions to your phone](#)

**via Great Western Hwy/Parramatta Rd 18 min**  
Fastest route, the usual traffic  
10.5 km  
[DETAILS](#)

**via Great Western Hwy/Parramatta Rd/A22 21 min**  
10.1 km  
[DETAILS](#)

**via A4 21 min**  
11.4 km  
[DETAILS](#)

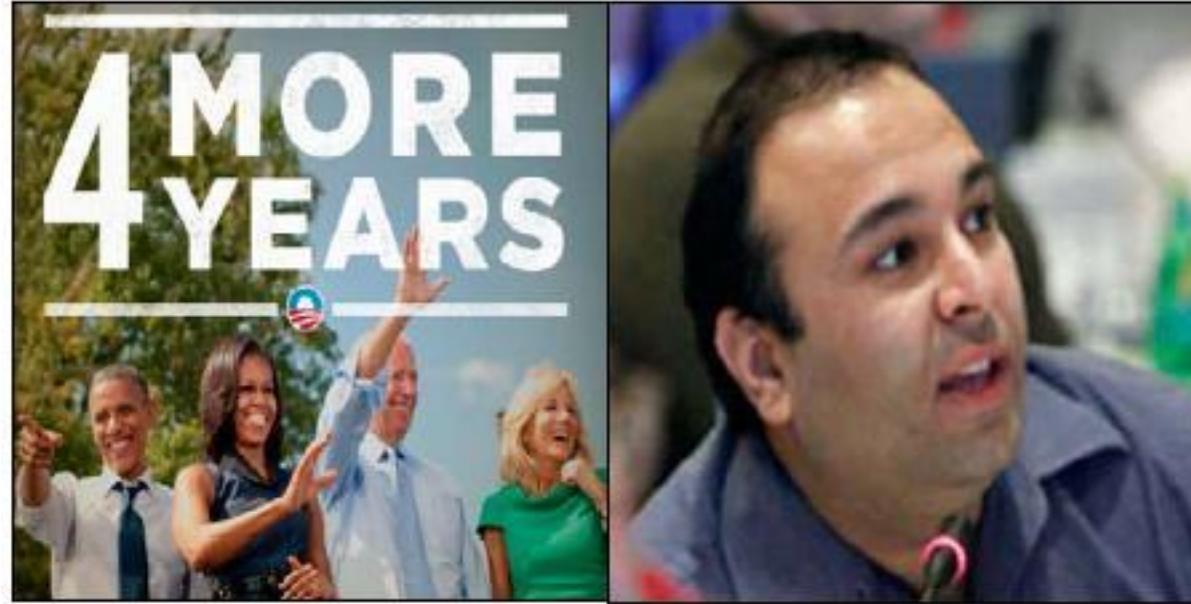


# Can RS be used for election?



# Data won the U.S. election

Data scientist Rayid Ghani helped persuade voters to reelect President Obama.



# Mining public opinions

 **Donald J. Trump**   
@realDonaldTrump

CHINA!

下午9:01 · 2020年5月29日 · Twitter for iPhone

14.2万 转推 8.1万 引用推文 75.8万 喜欢次数

...  
 **Donald J. Trump**  @realDonaldTrump · 2小时  
Progress!

 **NBC12 WWBT Richmond**  @NBC12 · 12小时  
Virginia Wesleyan University business professor and dean Paul Ewell wrote that anyone who chose Biden for president is "ignorant, anti-American and anti-Christian." nbc12.com/2020/11/14/va-...

回复 1.5万 转发 2.2万 喜欢 9.7万 分享

 **Donald J. Trump**  @realDonaldTrump · 3小时  
He only won in the eyes of the FAKE NEWS MEDIA. I concede NOTHING! We have a long way to go. This was a RIGGED ELECTION!

 This claim about election fraud is disputed

回复 8万 转发 9.6万 喜欢 31.3万 分享

# Improve poll ratings

 President Biden @POTUS  
United States government official

It was great to meet with you, @bts\_bighit. Thanks for all you're doing to raise awareness around the rise in anti-Asian hate crimes and discrimination.

I look forward to sharing more of our conversation soon.

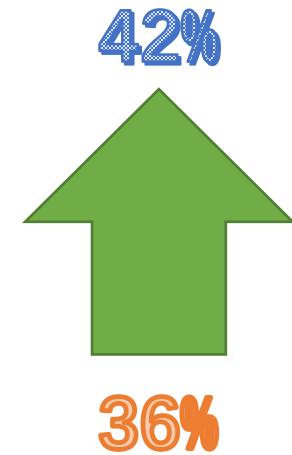


11:40 AM · Jun 1, 2022

Heart 1M Reply Share

Read 15.6K replies

- The tweet has attracted more than 500,000 retweets, along with over 1M likes.



<https://www.forbes.com/sites/nicholasreimann/2022/06/04/biden-posts-his-top-tweet-as-president-thanks-to-bts/?sh=41f7841274fb>

# Laws for Recommendation

 全国人民代表大会  
The National People's Congress of the People's Republic of China

首页 | 宪法 | 人大机构 | 票选书委员长 | 代表大会议 | 常委会会议 | 委员长会议 | 权威发布 | 立法 | 监督 | 代表  
对外交往 | 选举任免 | 法律研究 | 理论 | 机关工作 | 地方人大 | 图片 | 视频 | 直播 | 专题 | 资料库 | 国旗 | 国歌 | 国徽

当前位置：首页

**中华人民共和国数据安全法**  
(2021年6月10日第十三届全国人民代表大会常务委员会第二十九次会议通过)

来源：中国人大网 浏览字号：大 中 小 2021年06月10日 19:58:46

 全国人民代表大会  
The National People's Congress of the People's Republic of China

首页 | 宪法 | 人大机构 | 票选书委员长 | 代表大会议 | 常委会会议 | 委员长会议 | 权威发布 | 立法 | 监督 | 代表  
对外交往 | 选举任免 | 法律研究 | 理论 | 机关工作 | 地方人大 | 图片 | 视频 | 直播 | 专题 | 资料库 | 国旗 | 国歌 | 国徽

当前位置：首页

**中华人民共和国个人信息保护法**  
(2021年8月20日第十三届全国人民代表大会常务委员会第三十次会议通过)

来源：中国人大网 浏览字号：大 中 小 2021年08月20日 16:53:44

**目录**

- 第一章 总 则
- 第二章 数据安全与发展
- 第三章 数据安全制度
- 第四章 数据安全保护义务
- 第五章 政务数据安全与开放
- 第六章 法律责任
- 第七章 附 则

**目录**

- 第一章 总 则
- 第二章 个人信息处理规则
- 第一节 一般规定
- 第二节 敏感个人信息的处理规则
- 第三节 国家机关处理个人信息的特别规定
- 第三章 个人信息跨境提供的规则
- 第四章 个人在个人信息处理活动中的权利
- 第五章 个人信息处理者的义务
- 第六章 履行个人信息保护职责的部门
- 第七章 法律责任
- 第八章 附 则

 国家市场监督管理总局  
State Administration for Market Regulation

请输入要查询的内容

首页 机构 新闻 政务 服务 互动 专题

你的位置:首页 > 新闻 > 媒体聚焦

**互联网信息服务算法推荐管理规定**

发布时间: 2022-01-04 12:58 信息来源: 中国网信网

国家互联网信息办公室 中华人民共和国工业和信息化部 中华人民共和国公安部 国家市场监督管理总局

令 第9号

《互联网信息服务算法推荐管理规定》已经2021年11月16日国家互联网信息办公室2021年第20次室务会议审议通过，并经工业和信息化部、公安部、国家市场监督管理总局同意，现予公布，自2022年3月1日起施行。

国家互联网信息办公室主任 庄荣文  
工业和信息化部部长 肖亚庆  
公安部部长 赵克志  
国家市场监督管理总局局长 张 工  
2021年12月31日

**互联网信息服务算法推荐管理规定**

# Content

- Recommendation age
- A case study of recommender systems
- The 3C principle to build recommender systems
- Classic recommender systems
- Evaluation metrics for recommender systems

# Item domain recommendation

- Item types
  - E.g., flights, stays, cars
- Item relations:
  - E.g., destination (flight), location (hotel)

The screenshot shows the Expedia website interface for travel search. The top navigation bar includes links for English, List your property, Support, Trips, and Sign in. Below the bar, there are tabs for Stays, Flights, Cars, Packages, Things to do, and More travel. The 'Packages' tab is currently selected. A red box highlights the 'Going to' field, which is set to 'Sydney (and vicinity), New South Wales, Aust...'. Another red box highlights the 'Travelers' section, indicating '1 room, 3 travelers'. The search form also includes fields for 'Leaving from' (Shanghai, China (PVG-Pudong Intl.)), 'Preferred class' (Economy), 'Departing' (Dec 13), and 'Returning' (Dec 21). There are also checkboxes for 'Direct flights only' and 'I only need accommodations for part of my trip'. A large blue 'Search' button is centered below the form. Below the search bar, the breadcrumb navigation shows 'Choose your stay > Choose departing flight > Choose returning flight > Review your trip'. The main content area displays a list of accommodation options. At the top of this list is a message: 'Trip prices include roundtrip flight + stay, taxes, and fees. What we are paid impacts our sort order.' To the right is a dropdown menu labeled 'Sort by Recommended'. The first listing is for 'Ovolo 1888 Darling Harbour Pyrmont', which is marked as an 'Ad'. It features a photo of a modern hotel room, a rating of 4.7/5 (Exceptional), 997 reviews, and a price of '\$11,081 per person includes flight + stay'. The second listing is for 'Swissotel Sydney', located in the Sydney Central Business District, marked as an 'Unreal Deal'. It features a photo of an outdoor pool area, a rating of 4.5/5 (Wonderful), 1,278 reviews, and a price of '\$11,118 per person includes flight + stay'. Other filters and search options are visible on the left side of the results.

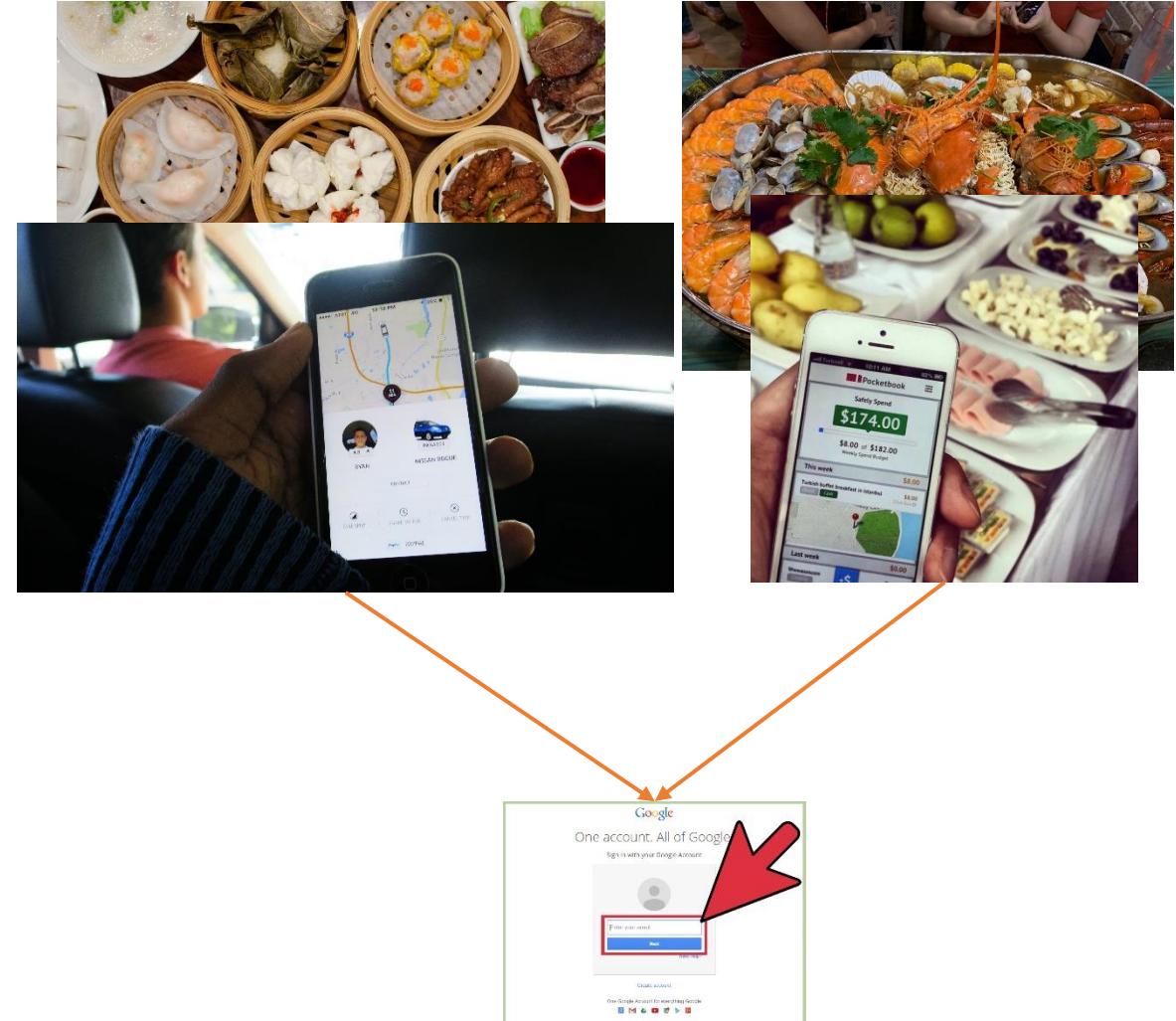
# User domain recommendation

- User groups
  - E.g., family vs. colleagues for stays
- Social relations:
  - E.g., local food recommendations according to your local friends



# Spatial domain recommendation

- Geographic points
  - E.g., Chinese food restaurants (Hongkong), Seafood restaurants (Sydney)
- Systems/Sites:
  - E.g., Uber (google account), Eats Apps (google account)



# Temporal domain recommendation

- Sessions:
  - E.g. a shopping transaction may help travel package recommendation
- Time period:
  - E.g., preferences of breakfast (morning) help to recommend dinner (evening)

*Recent transaction*



*Travel package*



*Breakfast*

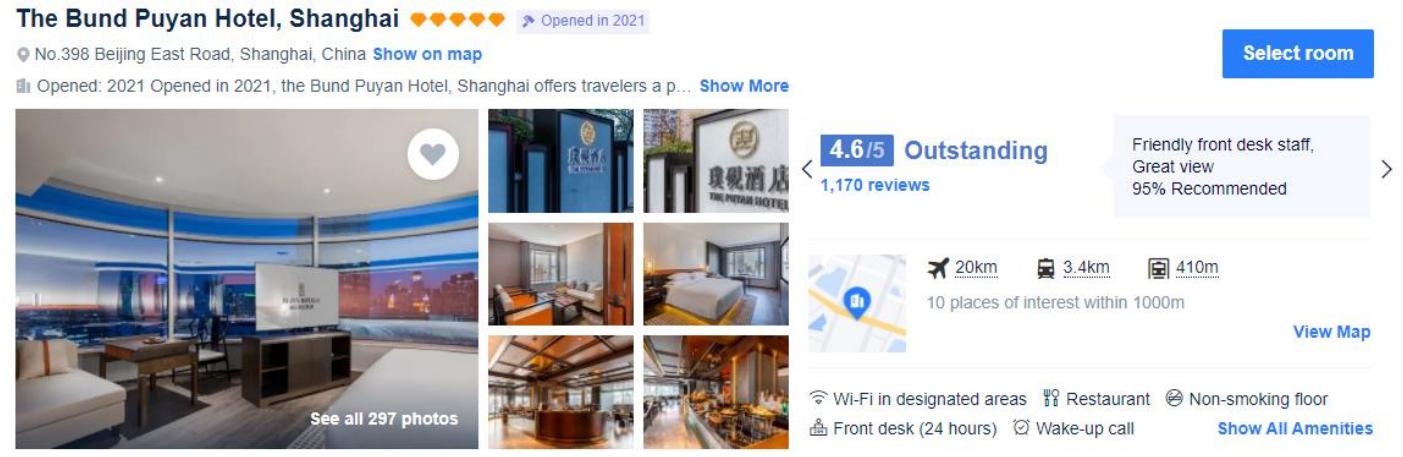


*Dinner*



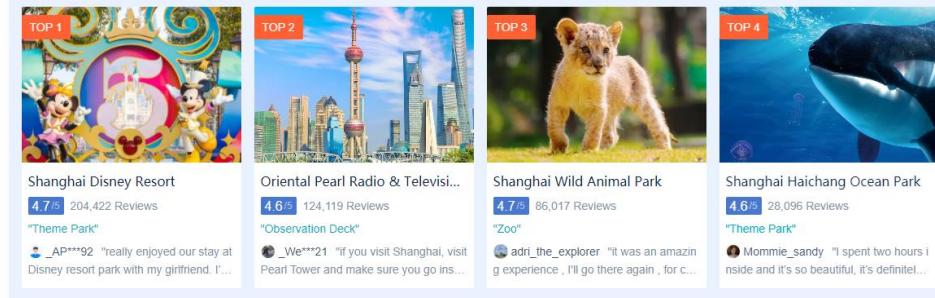
# Data domain recommendation

- Modalities:
  - E.g. rating, description, photos, geo information
- Distributions:
  - Ratings have different distributions for different domains, e.g., food vs stays
  - The content of image distribution is also different, e.g., color distributions between food and stays.

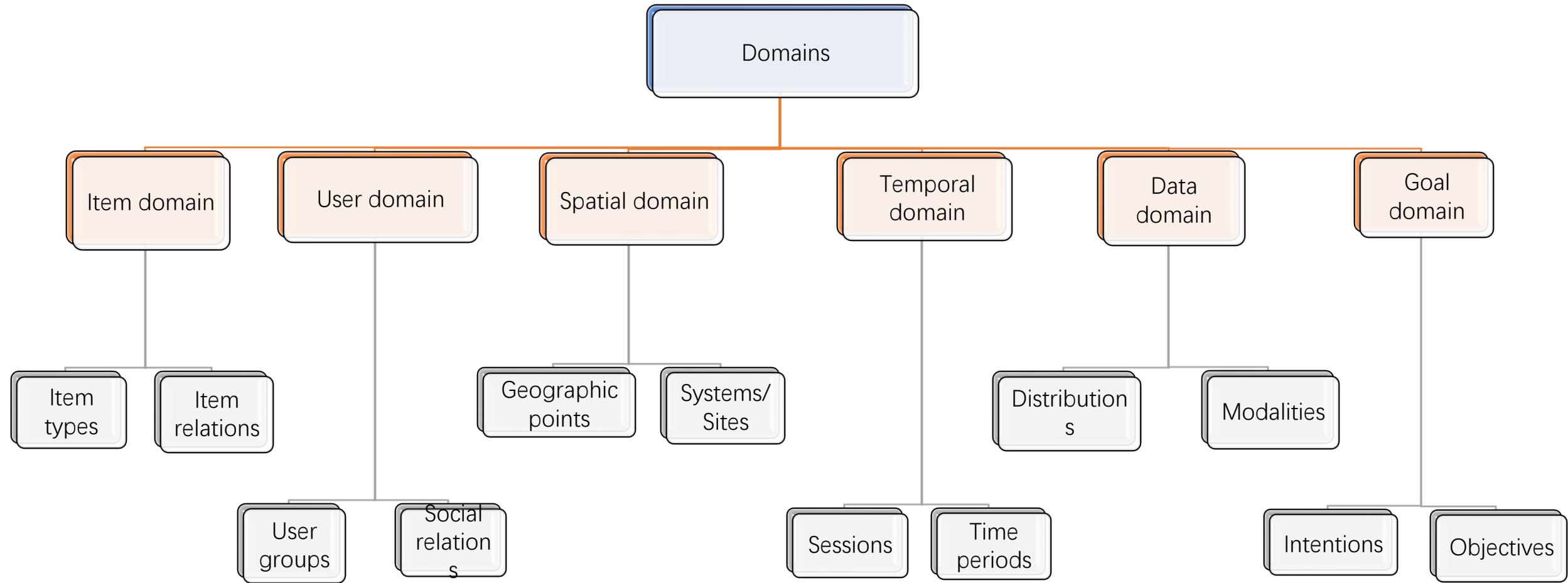


# Goal domain recommendation

- Intentions:
  - A travel often consists of multiple intentions:
  - E.g. visiting attractions, enjoy local food.
- Objectives:
  - Multiple criteria:
    - E.g. location, cleanliness, services
  - Multiple tasks:
    - E.g. booking the cheapest flight and a sea view hotel



# Modeling RS from different domains



# Content

- Recommendation age
- A case study of recommender systems
- The 3C principle to build recommender systems
- Classic recommender systems
- Evaluation metrics for recommender systems

# 3C: Complement, Composite, Context

- **Complement**
  - To provide a complete view for recommendation by leveraging complement information over multiple domains.
- **Composite**
  - To provide comprehensive views for recommendation by integrating composite information over multiple domains.
- **Context**
  - To provide a conscious view for recommendation by considering context information over multiple domains.

# Case study: recommendations for a travel



Rating summary

Location	5 circles
Sleep Quality	5 circles
Rooms	5 circles
Service	5 circles
Value	5 circles
Cleanliness	5 circles

Expedia

Stays Flights Cars Packages Things to do More travel ▾

Leaving from Shanghai, China (PVG-Pudong Intl.) Going to Sydney (and vicinity), New South Wales, Australia

Travelers 1 room, 3 travelers

Preferred class Economy Departing Dec 13 Returning Dec 21

Direct flights only  
 I only need accommodations for part of my trip

**Search**

Choose your stay > Choose departing flight > Choose returning flight > Review your trip

Trip prices include roundtrip flight + stay, taxes, and fees.  
What we are paid impacts our sort order ⓘ

Sort by Recommended

Stay flexible with free cancellation for the property  
We recommend booking a stay with free cancellation for the property in case your plans change.

Filter properties by free cancellation

Ovolo 1888 Darling Harbour Pyrmont

Book now to receive Ovolo Extras  
Free mini bar replenished daily + Wi-Fi included, plus Self Service Laundry inclusive and also Gym Access at no extra cost

4.7/5 Exceptional (997 reviews)

\$11,081 per person includes flight + stay

Swissotel Sydney Sydney Central Business District

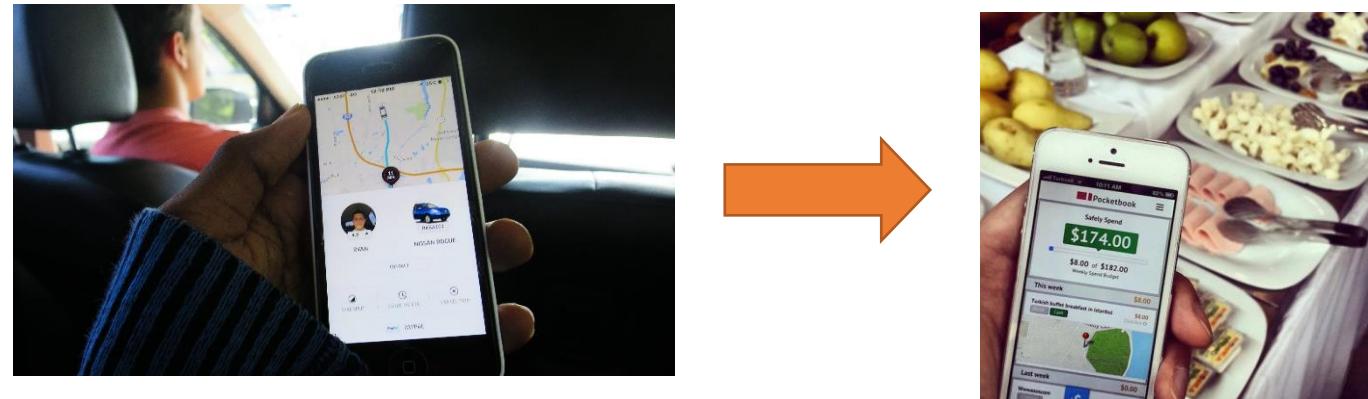
Free Cancellation Property

4.5/5 Wonderful (1,278 reviews)

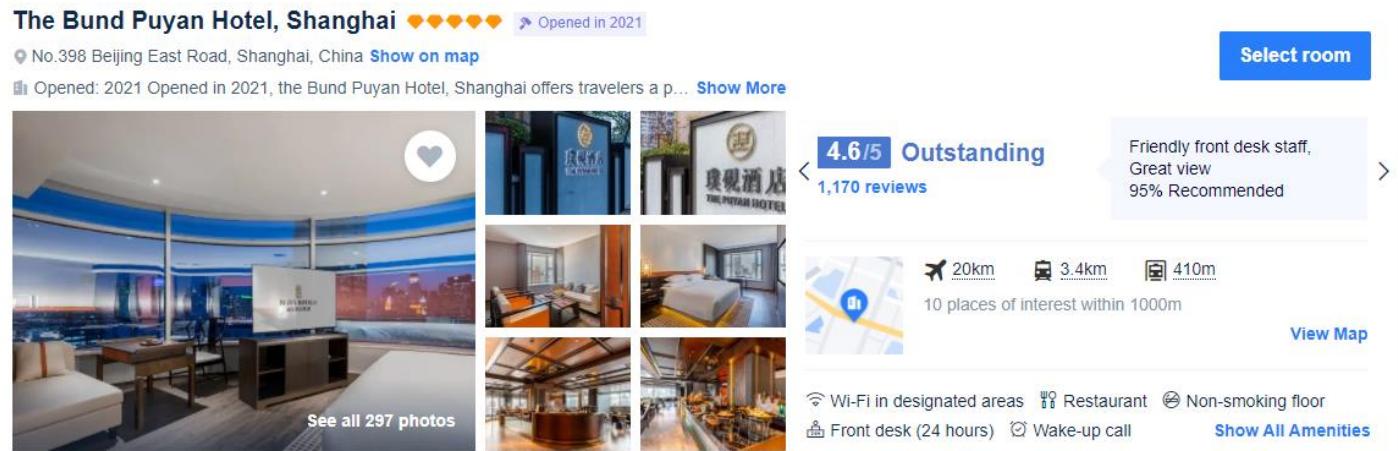
\$11,118 per person includes flight + stay

# Examples: complement information

- Transfer Uber histories for local restaurant recommendation

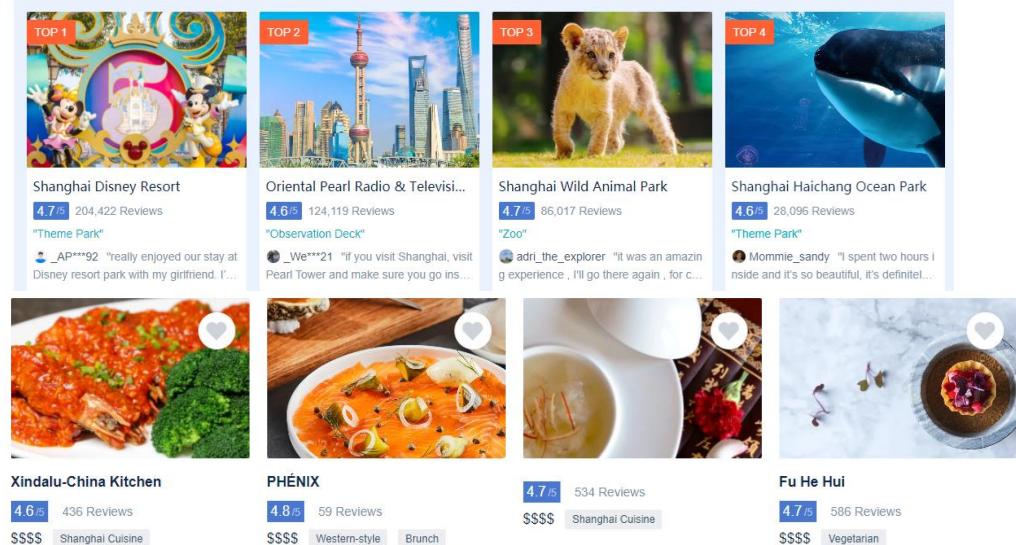


- Multimodal modeling for describing a hotel



# Examples: composite information

- A travel consists of multiple requirements, including attraction, food, and living



- A recommendation for a hotel needs to comprehensively consider multiple aspects



Rating summary

Location



Sleep Quality



Rooms



Service



Value



Cleanliness



# Examples: context information

- A recommendation needs to consider the time and location

*Breakfast*



*Dinner*



- A recommendation needs to consider the preferences of all members in a user group



# The way to the metaverse of RSs



Xindalu-China Kitchen

4.6 ⚡ 436 Reviews  
\$\$\$\$ Shanghai Cuisine



PHENIX

4.8 ⚡ 59 Reviews  
\$\$\$\$ Western-style Brunch



Fu He Hui

4.7 ⚡ 534 Reviews  
\$\$\$\$ Shanghai Cuisine



Fu He Hui

4.7 ⚡ 586 Reviews  
\$\$\$\$ Vegetarian



Shanghai Disney Resort

4.7 ⚡ 204,422 Reviews  
"Theme Park"

AP\*\*\*92 "really enjoyed our stay at Disney resort park with my girlfriend. I..."



Oriental Pearl Radio & Televisi...

4.6 ⚡ 124,119 Reviews  
"Observation Deck"

We\*\*\*21 "if you visit Shanghai, visit Pearl Tower and make sure you go ins..."



Shanghai Wild Animal Park

4.7 ⚡ 86,017 Reviews  
"Zoo"

adri\_the\_explorer "it was an amazing experience , I'll go there again , for c..."



Shanghai Haichang Ocean Park

4.6 ⚡ 28,096 Reviews  
"Theme Park"

Mommie\_sandy "I spent two hours inside and it's so beautiful, it's definitel..."



Rating summary

Location



Sleep Quality



Rooms



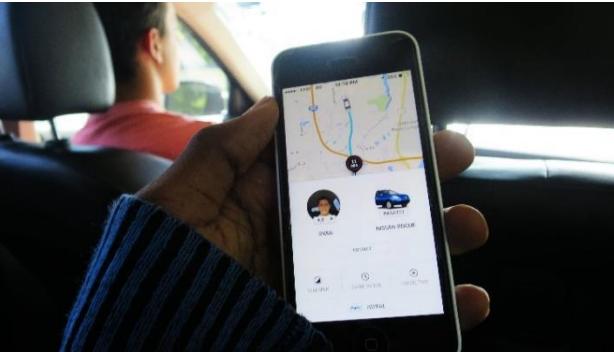
Service



Value



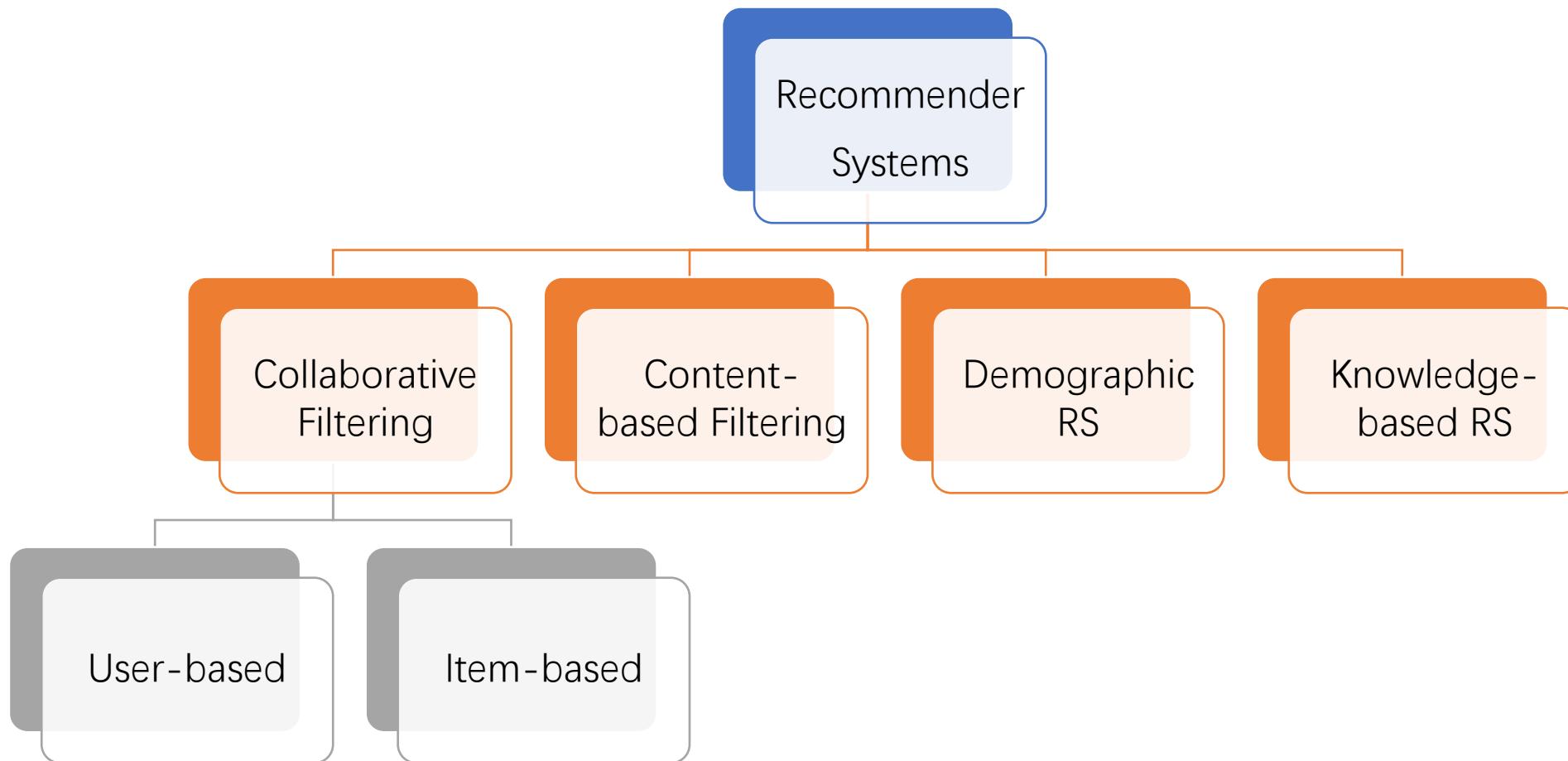
Cleanliness



# Content

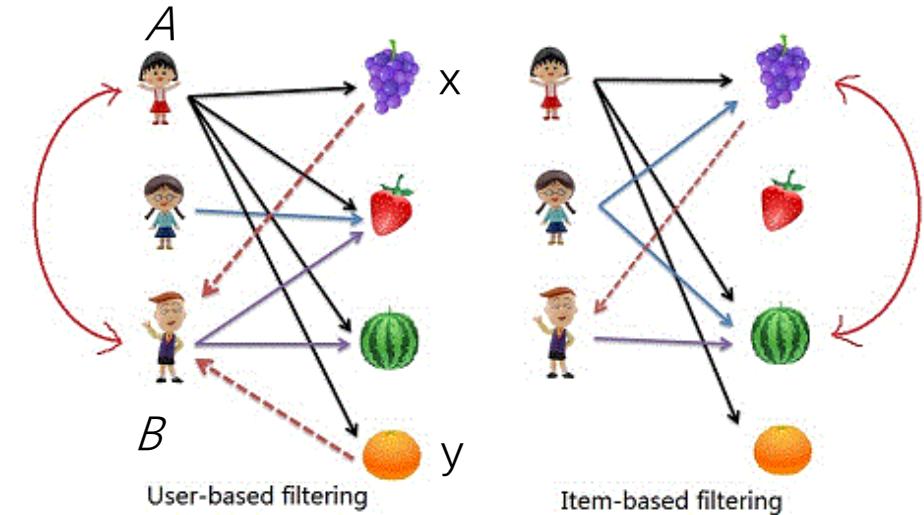
- Recommendation age
- A case study of recommender systems
- The 3C principle to build recommender systems
- Classic recommender systems
- Evaluation metrics for recommender systems

# Classic recommender systems



# Collaborative Filtering (CF)

- Intuition (user-based filtering): If user **A** related to user **B** and **A** bought x and y, then **B** bought x tend to buy y.
- Famous examples(item-based filtering): Amazon.com's recommender system
- Facebook, MySpace, LinkedIn use collaborative filtering to recommend new friends, groups, and other social connections.



# Memory-based CF

- The memory-based approach uses user rating data to compute the similarity between users or items.
- Typical examples of this approach are neighborhood-based CF and item-based/user-based top-N recommendations.
- User-based CF:

$$r_{\langle u, i \rangle} = \frac{1}{Z} \sum_{v \in \Gamma(u)} sim(u, v) r_{\langle v, i \rangle}$$

- Item-based CF:

$$r_{\langle u, i \rangle} = \frac{1}{Z} \sum_{j \in \Gamma(i)} sim(i, j) r_{\langle u, j \rangle}$$



# CF for friend recommendation

- Similarity (Jaccard)

- $s_{\langle u_1, u_2 \rangle} = \frac{|\{u_2, u_5, u_6\} \cap \{u_1, u_3, u_6\}|}{|\{u_2, u_5, u_6\} \cup \{u_1, u_3, u_6\}|} = \frac{1}{5}$
- $s_{\langle u_1, u_3 \rangle} = \frac{1}{4}, s_{\langle u_1, u_4 \rangle} = \frac{1}{4}$
- $s_{\langle u_1, u_5 \rangle} = \frac{1}{4}, s_{\langle u_1, u_6 \rangle} = \frac{2}{5}$

- Ranking

- $r_{\langle u_1, u_3 \rangle} = s_{\langle u_1, u_2 \rangle} * 1_{\langle u_2, u_3 \rangle} + s_{\langle u_1, u_3 \rangle} * 0_{\langle u_3, u_3 \rangle} + s_{\langle u_1, u_4 \rangle} * 1_{\langle u_4, u_3 \rangle} + s_{\langle u_1, u_5 \rangle} * 0_{\langle u_5, u_3 \rangle} + s_{\langle u_1, u_6 \rangle} * 0_{\langle u_6, u_3 \rangle} = \frac{9}{20}$
- $r_{\langle u_1, u_4 \rangle} = s_{\langle u_1, u_2 \rangle} * 0_{\langle u_2, u_4 \rangle} + s_{\langle u_1, u_3 \rangle} * 1_{\langle u_3, u_4 \rangle} + s_{\langle u_1, u_4 \rangle} * 0_{\langle u_4, u_4 \rangle} + s_{\langle u_1, u_5 \rangle} * 0_{\langle u_5, u_4 \rangle} + s_{\langle u_1, u_6 \rangle} * 1_{\langle u_6, u_4 \rangle} = \frac{13}{20}$

	$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$
$u_1$	0	1	0	0	1	1
$u_2$	1	0	1	0	0	1
$u_3$	0	1	0	1	0	0
$u_4$	0	0	1	0	0	1
$u_5$	1	0	0	0	0	1
$u_6$	1	1	0	1	1	0

# CF for rating prediction

- Similarity (Pearson correlation)

$$s_{\langle \textcolor{red}{u}_1, \textcolor{green}{u}_2 \rangle} = \frac{\sum_{i \in I(\textcolor{red}{u}_1, \textcolor{green}{u}_2)} (r_{\langle \textcolor{red}{u}_1, \textcolor{violet}{i} \rangle} - \bar{r}_{\textcolor{red}{u}_1})(r_{\langle \textcolor{green}{u}_2, \textcolor{violet}{i} \rangle} - \bar{r}_{\textcolor{green}{u}_2})}{\sqrt{\sum_{i \in I(\textcolor{red}{u}_1, \textcolor{green}{u}_2)} (r_{\langle \textcolor{red}{u}_1, \textcolor{violet}{i} \rangle} - \bar{r}_{\textcolor{red}{u}_1})^2} \sqrt{\sum_{i \in I(\textcolor{red}{u}_1, \textcolor{green}{u}_2)} (r_{\langle \textcolor{green}{u}_2, \textcolor{violet}{i} \rangle} - \bar{r}_{\textcolor{green}{u}_2})^2}}$$

- Rating Prediction

$$r_{\langle \textcolor{red}{u}, \textcolor{violet}{i} \rangle} = \bar{r}_{\textcolor{red}{u}} + \frac{\sum_{v \in \Gamma(u)} s_{\langle \textcolor{red}{u}_1, v \rangle} (r_{\langle v, \textcolor{violet}{i} \rangle} - \bar{r}_v)}{\sum_{v \in \Gamma(u)} s_{\langle \textcolor{red}{u}_1, v \rangle}}$$

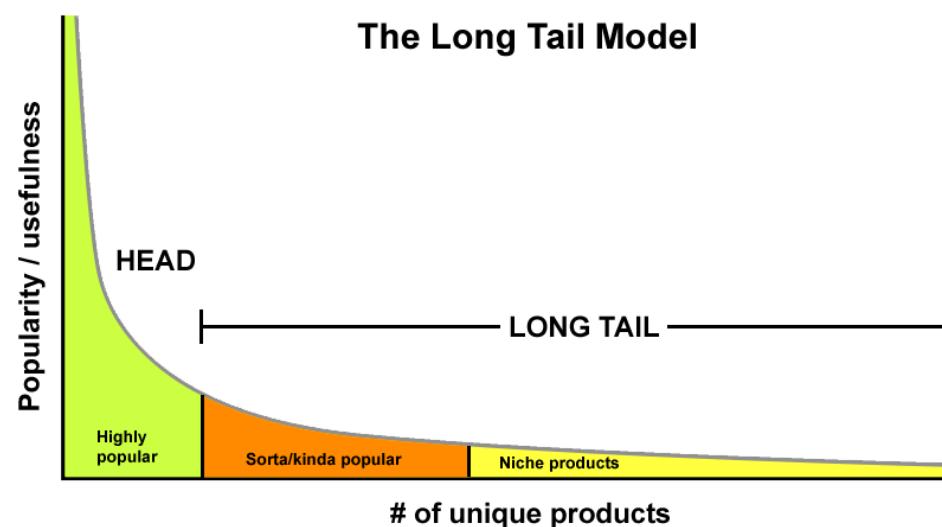
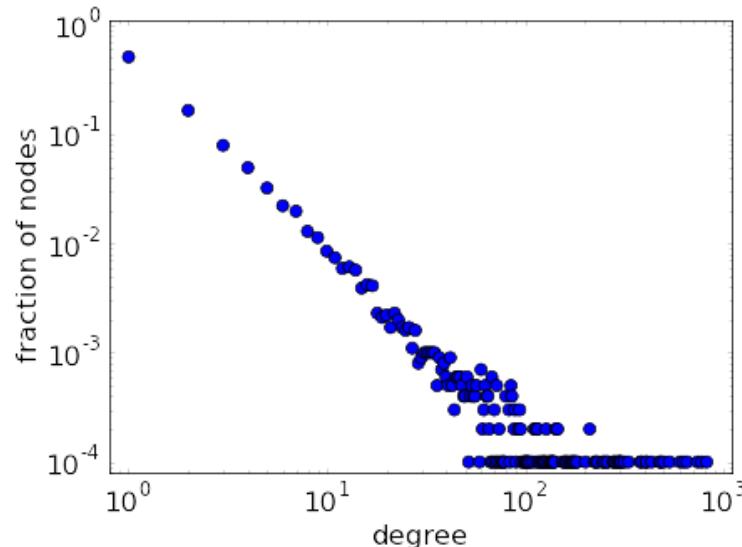
- Example:

- $\bar{r}_{\textcolor{red}{u}_1} = 4, \bar{r}_{\textcolor{red}{u}_2} = 3.5, \bar{r}_{\textcolor{red}{u}_3} = 4, \bar{r}_{\textcolor{red}{u}_4} = 5, \bar{r}_{\textcolor{red}{u}_5} = 1, \bar{r}_{\textcolor{red}{u}_6} = 3$
- $s_{\langle \textcolor{red}{u}_1, \textcolor{green}{u}_2 \rangle} = 0, s_{\langle \textcolor{red}{u}_1, \textcolor{green}{u}_3 \rangle} = 0.7071, s_{\langle \textcolor{red}{u}_1, \textcolor{green}{u}_4 \rangle} = 0$
- $s_{\langle \textcolor{red}{u}_1, \textcolor{green}{u}_5 \rangle} = 0, s_{\langle \textcolor{red}{u}_1, \textcolor{green}{u}_6 \rangle} = 1$
- $r_{\langle \textcolor{red}{u}_1, \textcolor{violet}{i}_1 \rangle} = 4 + \frac{0.7071 * (4 - 4) + 1 * (2 - 3)}{0.7071 + 1} = 4 - 0.5858 = 3.4142$

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$u_1$		5		4	3
$u_2$	3			4	
$u_3$	4	5	3	4	4
$u_4$				5	
$u_5$	1				
$u_6$	2	4		3	

# Data characteristics in recommender systems

- Power law or Long tail distribution
  - Data associated with the **majority** of users are **insufficient** and even **absent** in real world.
  - In most recommender systems, the **majority** of users/items only associated with very **few data** while only **minority** of users/items have **sufficient data**



# Challenges in collaborative filtering

- **Data Sparsity**

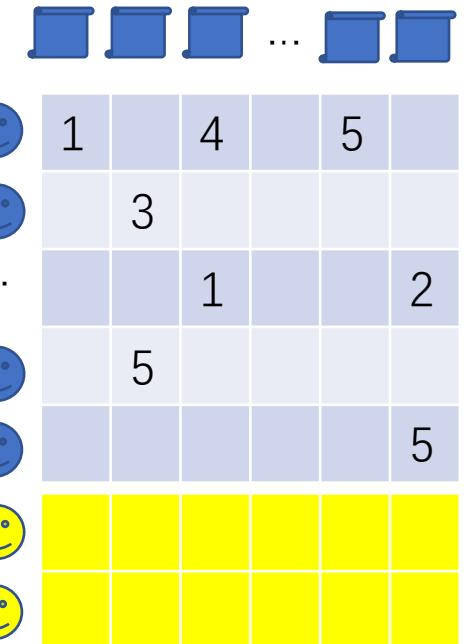
- In real-world recommender systems, the user-item matrix is very sparse.

- **Cold Start**

- When **new users or new items** are added, the system cannot recommend to these users and these items.

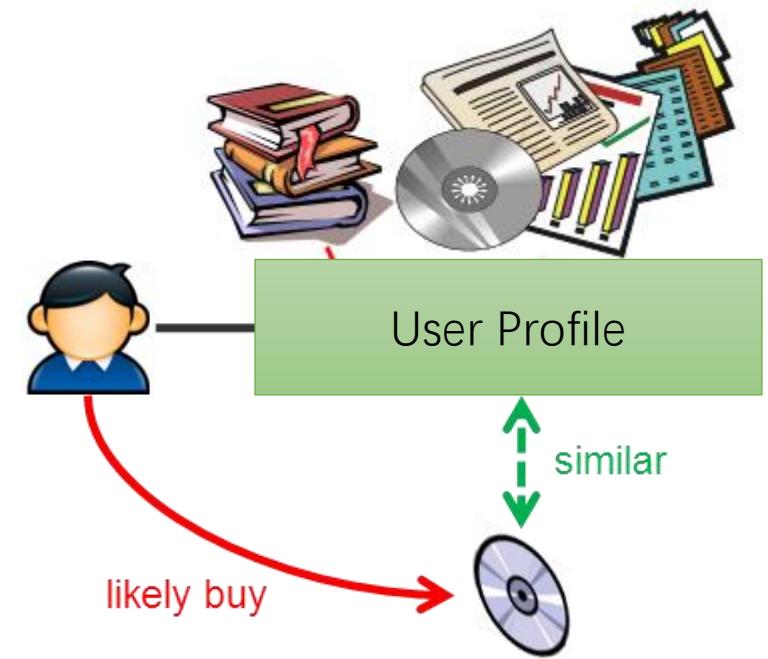
- **Scalability**

- There are **millions of users and products** in real systems.
  - Large amount of computation
  - Large storage



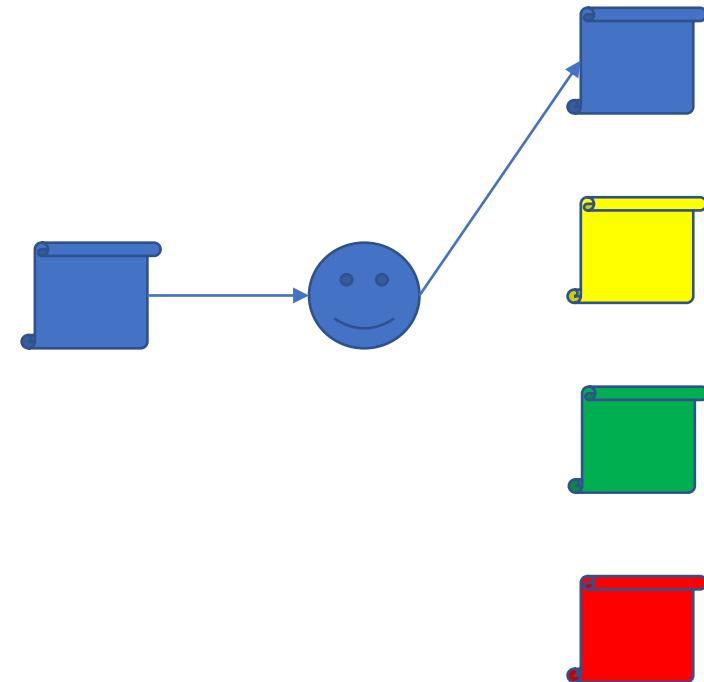
# Content-based Filtering (CBF)

- CBF is based on the features of items
  - Attributes of items
  - Description of items
  - Text of an article
- User profile is built with the features of historical items
- Recommend items according to user profile



# Challenges in content-based filtering

- **Limited Content Analysis**
  - System has a limited amount of information on its users or the content of its items.
- **Over-specialization**
  - The system can only recommend items that highly similar with user's profile, the user is limited to be recommended items similar to those already rated.



# Demographic RS

- The demographic RS provides suggestions based on the characteristics of the specific audience segments.
- This type of recommendation engine takes available user data (age, gender, location, etc.), classifies it into specific audience segments, and then puts in a bigger picture to fill the gaps in the data.
- **Pros:** Non-existence of cold start
- **Cons:** Non-personalized recommendations

Name	Gender	Occupation	Country	Age
John	M	Student	France	13
Paul	M	Doctor	France	34
Sarah	F	Student	USA	12
Mike	M	Teacher	France	27

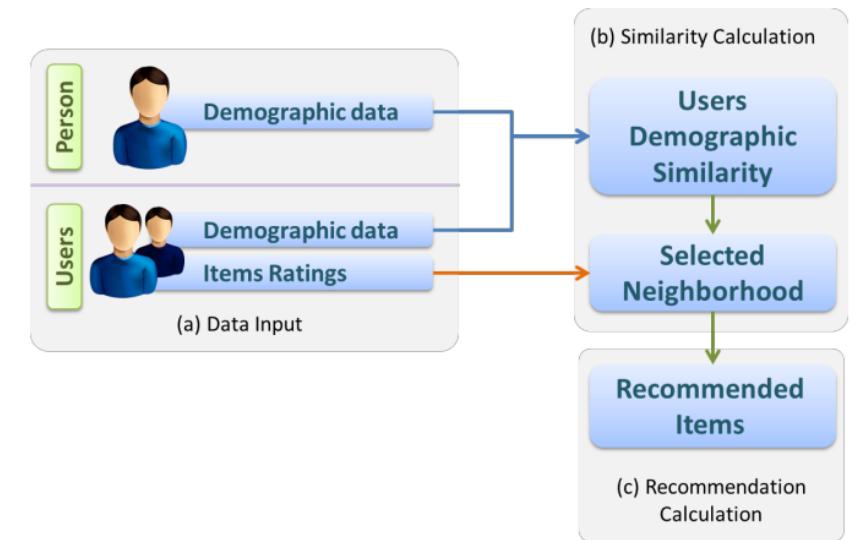
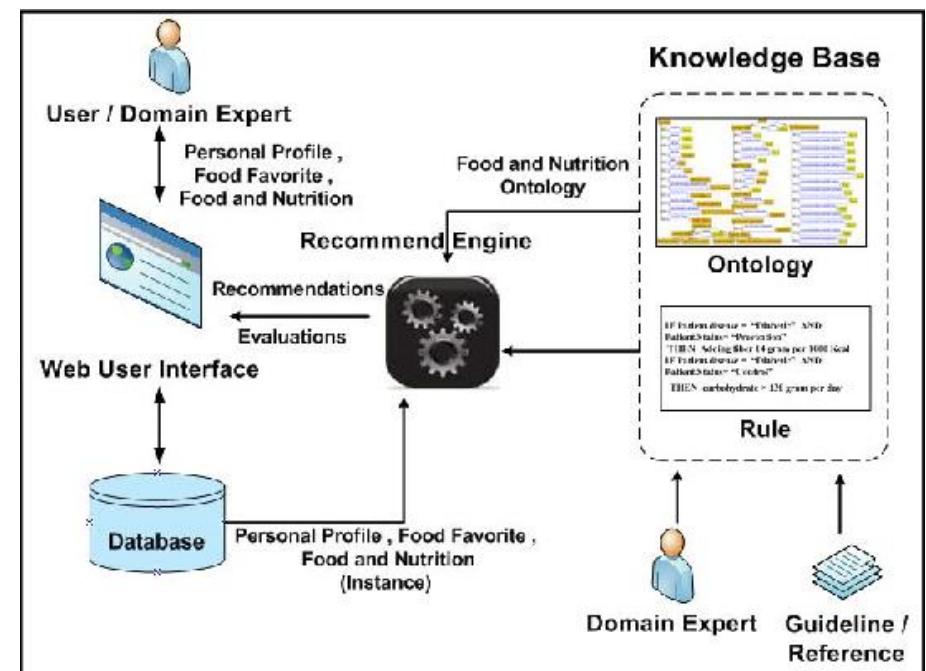


Fig. 2. Demographic-based approach for new users.

# Knowledge-based RS

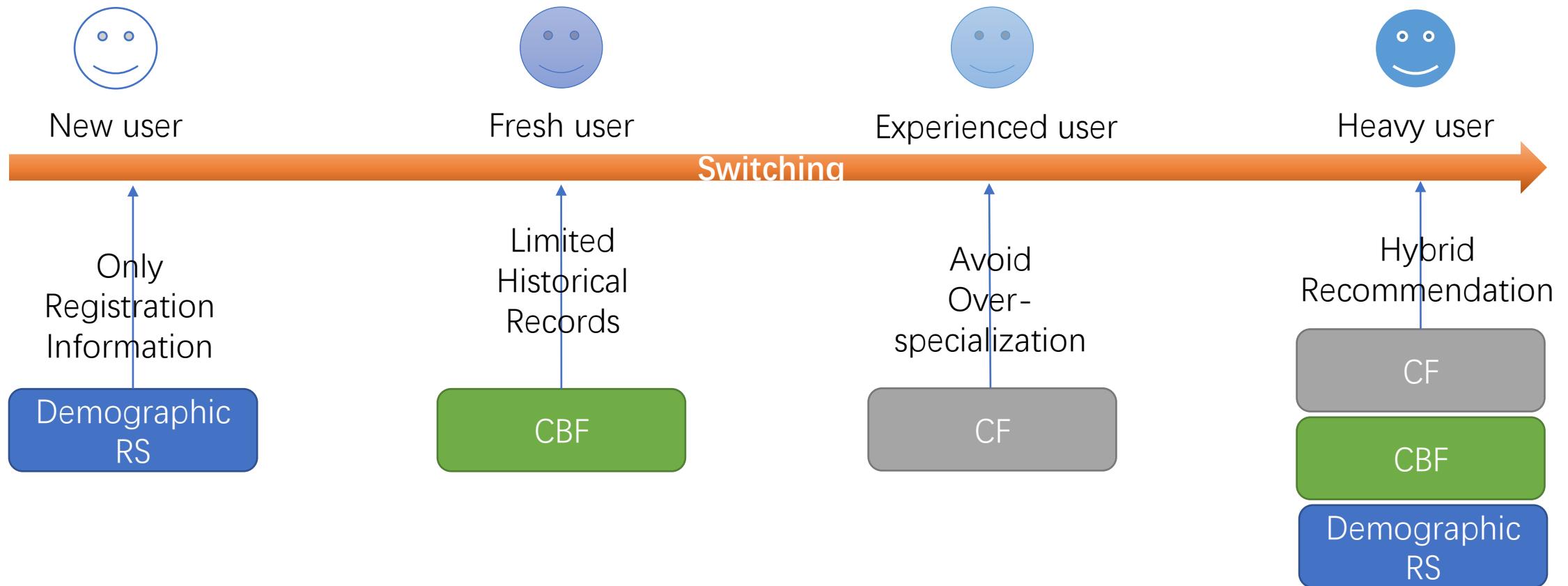
- Knowledge-based RS are based on explicit knowledge about the item assortment, user preferences, and recommendation criteria (i.e., which item should be recommended in which context).
- **Pros:** Non-existence of cold start
- **Cons:** Need to define recommendation knowledge in an explicit fashion



# Hybrid RS

- Most RSs use a hybrid approach, combining CF, CBF, and other approaches.
- Some hybridization techniques include:
  - **Weighted**: Combining the score of different recommendation components numerically.
  - **Switching**: Choosing among recommendation components and applying the selected one.
  - **Mixed**: Recommendations from different recommenders are presented together to give the recommendation.
  - **Feature Combination**: Features derived from different knowledge sources are combined together and given to a single recommendation algorithm.
  - **Feature Augmentation**: Computing a feature or set of features, which is then part of the input to the next technique.
  - **Cascade**: Recommenders are given strict priority, with the lower priority ones breaking ties in the scoring of the higher ones.
  - **Meta-level**: One recommendation technique is applied and produces some sort of model, which is then the input used by the next technique.

# Example: Hybrid RS

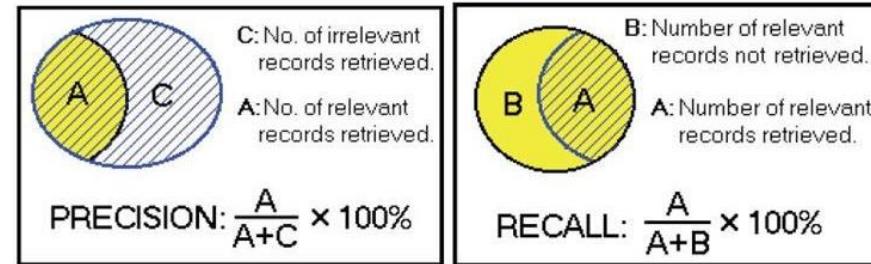


# Content

- Recommendation age
- A case study of recommender systems
- The 3C principle to build recommender systems
- Classic recommender systems
- Evaluation metrics for recommender systems

# Evaluation metric types

- Relevance-based evaluation



- Ranking-based evaluation



- Rating-based evaluation

	Item 1	Item 2	Item 3	...	Item n
User 1	2	3	?	...	5
User 2	?	4	3	...	?
User 3	3	2	?	...	3
...	...	...	...	...	...
User m	1	?	5	...	4

# Relevance-based evaluation

- Recommendation can be viewed as a special information retrieval task to recommend the items which relevant to users' preference.

- *Recall*
- *Precision*
- *Hit Ratio*

		Real Label	
		Positive	Negative
Predicted Label	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Precision =  $\frac{\sum \text{TP}}{\sum \text{TP} + \text{FP}}$

$\downarrow$

Recall =  $\frac{\sum \text{TP}}{\sum \text{TP} + \text{FN}}$

Accuracy =  $\frac{\sum \text{TP} + \text{TN}}{\sum \text{TP} + \text{FP} + \text{FN} + \text{TN}}$

# Recall@K

- The fraction of relevant items over all  $N$  relevant items

$$Recall@K = \frac{\sum_{k=1}^K TP_k}{N}$$



$$Recall@3 = \frac{2}{5}$$

# Precision@K

- The fraction of relevant items over top  $K$  recommended items

$$Precision@K = \frac{\sum_{k=1}^K TP_k}{K}$$

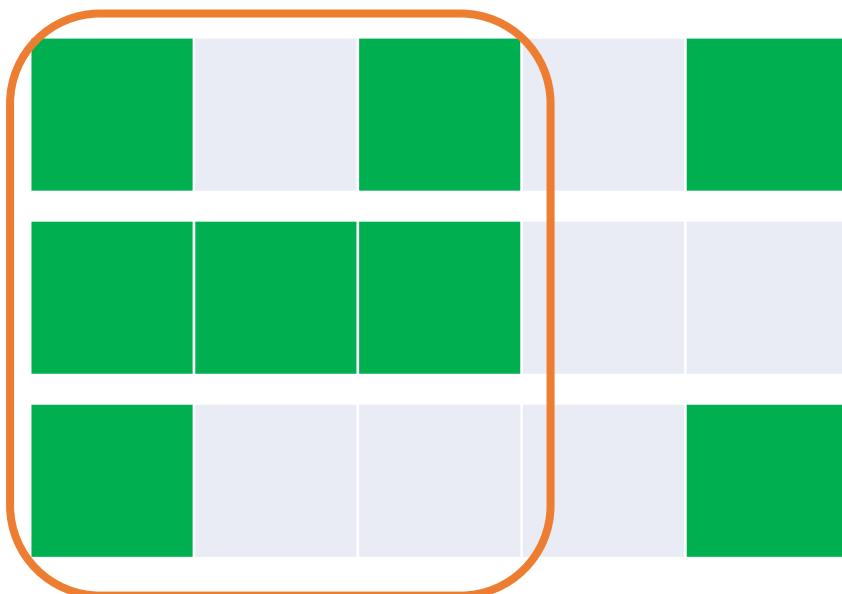


$$Recall@3 = \frac{2}{3}$$

# Hit Ratio (HR@K)

- The ratio of hit (predicted) true results w.r.t. the total ground truth results.

$$HR@K = \frac{\sum_{q=1}^Q \sum_{k=1}^K TP_{q,k}}{\sum_{q=1}^Q \min(N_q, K)}$$



$$HR@3 = \frac{2 + 3 + 1}{3 + 3 + 2} = \frac{6}{8}$$

# Ranking-based evaluation

- The most common way to assess the recommendation performance is to measure whether relevant items are placed in the top positions of a recommendation list.
  - ***MAP***: Mean Average Precision
  - ***MRR***: Mean Reciprocal Rank
  - ***nDCG***: Normalized Discounted Cumulative Gain

# AP@K

- Average Precision (AP) is a ranked precision metric that places emphasis on highly ranked correct predictions (hits).
- AP@K is the average result over Precision@1~K, which is defined as

$$AP@K = \frac{\sum_{k=1}^K relevant(k) * Precision@k}{\min(K, N)}$$

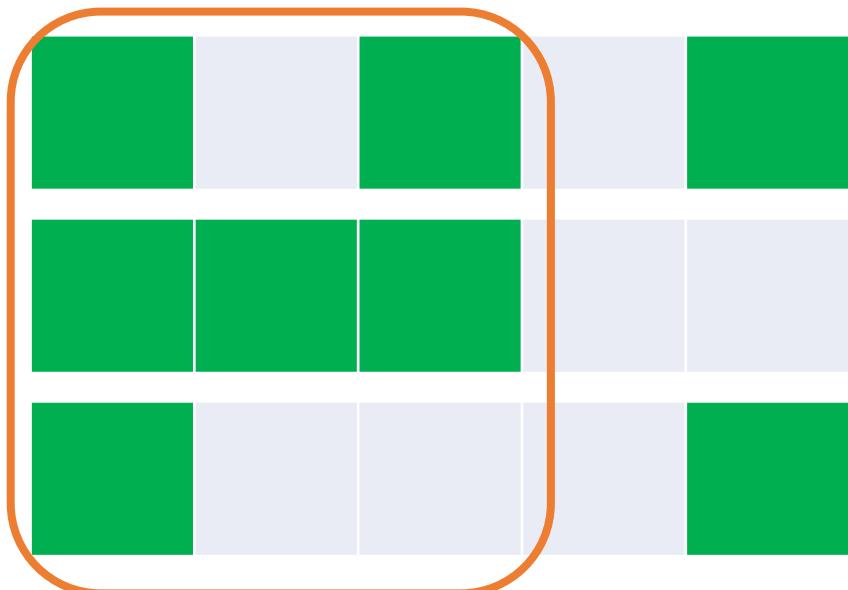


$$AP@4 = \frac{1 + 0 + \frac{2}{3} + 0}{\min(2, 4)} = \frac{5}{6}$$

# MAP@K

- Mean Average Precision (MAP) is the mean of AP over all the recommended lists:

$$MAR@K = \frac{\sum_{q=1}^Q AP_q @ K}{Q}$$



$$AP_1 @ 3 = \frac{\left(1 + \frac{2}{3}\right)}{3} = \frac{5}{9}$$

$$AP_2 @ 3 = \frac{(1 + 1 + 1)}{3} = 1$$

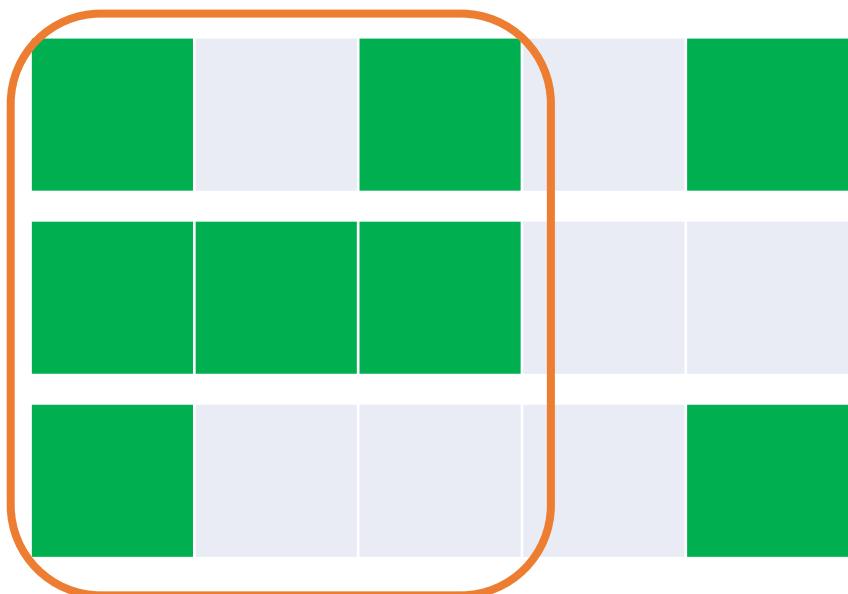
$$AP_3 @ 3 = \frac{1}{2}$$

$$MAR@3 = \frac{1}{3} \left( \frac{5}{9} + 1 + \frac{1}{2} \right)$$

# MRR@K

- Mean Reciprocal Rank evaluates any process that produces a list of possible responses to  $Q$  testing users, ordered by probability of correctness., which is defined as:

$$RR@K = \sum_{k=1}^K \frac{relevant(k)}{k}$$
$$MRR@K = \frac{\sum_{q=1}^Q RR_q @ K}{Q}$$



$$RR_1 @ 3 = \frac{1}{1} + \frac{1}{3} = \frac{4}{3}$$

$$RR_2 @ 3 = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} = \frac{4}{3}$$

$$RR_3 @ 3 = \frac{1}{1}$$

$$MRR@3 = \frac{1}{3} \left( \frac{4}{3} + \frac{4}{3} + 1 \right)$$

# nDCG@K

- Normalized Discounted Cumulative Gain is a measure of ranking quality which places strong emphasis on relevant items

$$nDCG@K = \frac{DCG@K}{IDCG@K}$$

- where

$$DCG@K = \sum_{i=1}^K \frac{2^{rel(i)} - 1}{\log_2(i + 1)}, \quad IDCG@K = \sum_{i=1}^K \frac{1}{\log_2(i + 1)}$$



$$DCG@4 = 1 + 0 + \frac{1}{\log_2(3 + 1)} + 0 = \frac{3}{2}$$
$$IDCG@K = \log_2 2 + \log_2 3 + \log_2 4 + \log_2 5$$

# Rating-based evaluations

- Rating-based evaluation aims to evaluate the prediction error between the true rating values and the predicted ones.
- Two commonly used rating-based evaluation metrics:
  - **MAE**: Mean Absolute Error
  - **RMSE**: Root Mean Squared Error

	Item 1	Item 2	Item 3	...	Item n
User 1	2	3	?	...	5
User 2	?	4	3	...	?
User 3	3	2	?	...	3
...	...	...	...	...	...
User m	1	?	5	...	4

# MAE & RMSE

- Metrics measuring error rate
  - **Mean Absolute Error (MAE)** computes the deviation between predicted ratings and actual ratings

$$MAE = \frac{\sum_{i=1}^N |r_i - p_i|}{N}$$

- **Root Mean Square Error (RMSE)** is similar to MAE, but places more emphasis on larger deviation

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (r_i - p_i)^2}{N}}$$

# Rating prediction: an example

# ID	User ID	Movie ID	Rating ( $r_i$ )	Prediction ( $p_i$ )	$ r_i - p_i $	$(r_i - p)^2$
1	1	134	5	4.5	0.5	0.25
2	1	238	4	5	1	1
3	1	312	5	5	0	0
4	2	134	3	5	2	4
5	2	767	5	4.5	0.5	0.25
6	3	68	4	4.1	0.1	0.01
7	3	212	4	3.9	0.1	0.01
8	3	238	3	3	0	0
9	4	68	4	4.2	0.2	0.04
10	4	112	5	4.8	0.2	0.04
Sum					4.6	5.6

$$\text{MAE} = 4.6/10 = 0.46$$

$$\text{RMSE} = \sqrt{5.6/10}$$

# Summary

- We have entered recommendation age
- There are more and more emerging needs for recommender systems
- Different recommendation approaches have their pros and cons
- The recommendation results should be evaluated with diverse metrics

# Thinking

- Try to find recommender systems in your daily life.
- Are the recommender systems only used in e-commerce?
- According to you, what are the difficulties to build recommend systems?

# Assignment

## MovieLens 100K Dataset

MovieLens 100K movie ratings. Stable benchmark dataset. 100,000 ratings from 1000 users on 1700 movies.

Task: Apply memory-based CF to predict movie ratings.

- Evaluation metrics: MAE, RMSE

Experimental dataset: <https://grouplens.org/datasets/movielens/100k/>

- Training set: ua.base
- Testing set: ua.test