

推荐系统概述

任务、特征与方法

(Recommender System)

徐悦牲(Yueshen Xu)

ysxu@xidian.edu.cn / xuyueshen@163.com

计算机科学与技术学院

智能软件与系统新技术研究所



西安电子科技大学
XIDIAN UNIVERSITY

□ 推荐系统 (Recommender System)

■ 背景、示例与应用

■ 推荐系统要素

■ 常用推荐方法

- 基于协同过滤的推荐方法
- 基于内容的推荐方法
- 基于标签的推荐方法
- 基于网络的推荐方法
- 基于上下文的推荐方法

■ 推荐方法测试

背景、示例与应用

→ 背景



西安电子科技大学
XIDIAN UNIVERSITY

□ 信息检索引擎的发展

- 信息检索引擎 != 搜索引擎

门户网站



搜索引擎



有合并的趋势：
垂直搜索引擎



背景、示例与应用

→ 背景



西安电子科技大学
XIDIAN UNIVERSITY

□ 这些都不够

■ 信息过载；信息爆炸

- 待搜索信息本身快速增长，寻找目标信息越发困难



■ 搜索展现力不够强

- 受页面与排版的限制

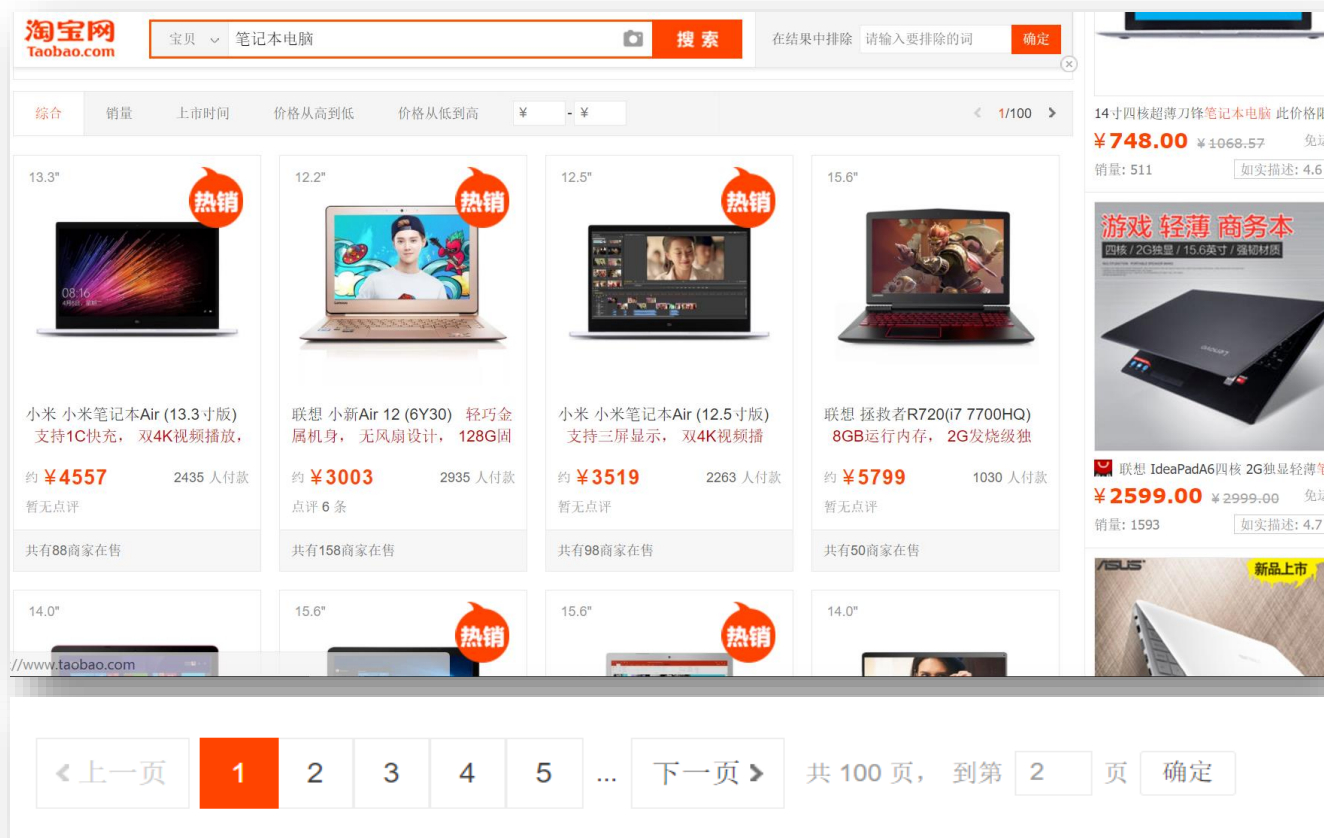


背景、示例与应用

→ 背景



西安电子科技大学
XIDIAN UNIVERSITY



有限的展示位
有限的页面
有限的条目

→ 你会点几次“下一页”？

背景、示例与应用

→ 背景



西安电子科技大学
XIDIAN UNIVERSITY

- 搜索展示能力的限制易导致马太效应
 - 强者越强，弱者越弱
 - 内容方（卖方）会通过其它途径弥补
 - 新/小网站 → SEO (搜索引擎优化), 提高排名
 - 盈利性网站（民营医院） → 广告, 竞价排名
 - 新/小卖家 → 广告, 刷单
- 但仍以展示“主流”结果为主，缺乏个性化

■ 用户自身个性化增强

- 世界越发多元化，改革开放四十年，经济发展，个性释放
- 普通人有了追求自我的机会与条件

背景、示例与应用

→ 背景



西安电子科技大学
XIDIAN UNIVERSITY

➤ 个性化的时代

- 文青，文艺范，小清新，欧美风，中国风，治愈系...
- 追星，追剧，追品牌...
- 极简主义，素食主义，自然主义，生活主义...
- 总之，人们不想完全从众



背景、示例与应用

→ 背景



西安电子科技大学
XIDIAN UNIVERSITY

■ 用户需求不能完全由关键词表达

- 用户也不完全清楚自己要搜索的内容
- 用户清楚，但使用关键字不当
- 待搜索内容不易使用关键词表达清楚
 - 太长，太具体，太多条件
- 搜索引擎的补救措施



背景、示例与应用

→ 背景



西安电子科技大学
XIDIAN UNIVERSITY

□ 推荐系统

- 别名：Query-free Information Retrieval

- v.s. 搜索引擎

- 搜索引擎：用户明晰自己想要什么，且知道如何表达需求
- 推荐系统：用户大概知道想要什么，但没有明确的表达

- 示例：Java的相关教材

- **使用门户网站**（当当/新华书店）：教材 → 大学教材 → 计算机类 → 编程类 → Java → 找
- **使用搜索引擎**(Google/Bing/百度)：Java教材
- **使用推荐系统**：Java从入门到精通，Java编程思想，Java程序设计，Java Web程序设计

背景、示例与应用

→ 示例与应用



西安电子科技大学
XIDIAN UNIVERSITY

□ 对推荐系统依赖性较强的场景

电子商务网站



在线社交网络



在线内容站点



今日头条

在线社区网络



在线综合类网络



背景、示例与应用

→ 示例与应用



西安电子科技大学
XIDIAN UNIVERSITY

□ 常见推荐形式

猜用户喜欢

猜你喜欢 根据你的喜好精心为你推荐!

排行榜 专属你的购物指南

大学教材 | 手机通讯 | 笔记本 | 路由器 | 清洁工具

(SUPOR) 电磁 C8E15-210 .00



外科学 (第3版 供8年制及7年制 "5



内科学(第8版) 葛均波、徐永健/本



高等数学 下 (第7版) /十二五普通



全国中医药行业高等教育 "十二



大学数学学习辅导丛书: 线性代数附



国外电子与通信教材系列: 半导体物

相关性推荐/买了又买/看了又看

与您浏览过的商品相关的推荐



¥ 73.10



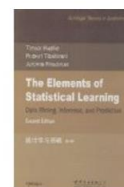
¥ 38.70



¥ 53.50



¥ 49.00



¥ 93.40

喜欢这部电影的人也喜欢 ·····



速度与激情7



湄公河行动



碟中谍5: 神秘国度



金蝉脱壳



谍影重重5

背景、示例与应用

→ 示例与应用



西安电子科技大学
XIDIAN UNIVERSITY

□ 常见推荐形式

精品推荐/热门推荐



火力推荐



关联互补



自由搭配



背景、示例与应用

→ 示例与应用



西安电子科技大学
XIDIAN UNIVERSITY

□ 常见推荐对象

■ 各种商品(广泛的概念, 一切可交易的物品)

- 实物 (大到游艇, 中到汽车, 小到书籍, 微到螺丝钉)、旅游线路、售后服务、电信套餐、App

■ (多媒体) 内容

- 视频、音乐、图片、标签、评论、网页、回答、节目

■ 关联关系

- 朋友、大V、偶像、小组、圈子、关注的人

■ 线下消费

- 各种形式的店, 比如餐馆、饮品店等

背景、示例与应用

→ 示例与应用



西安电子科技大学
XIDIAN UNIVERSITY

□ 推荐系统的成功案例

■ 亚马逊：主页即推荐

- 35%的销售额与推荐系统有关

■ Hulu

- 借助推荐系统快速崛起

■ 雅虎新闻

- 每个人看到的新闻都不一样 → 雅虎没了，但雅虎新闻还在

■ 今日头条：完全建立在新闻推荐系统之上

- 口号：“你关心的，才是头条”

■ Netflix：推荐即流量

- 利用推荐系统拍出了《纸牌屋》
- 推荐系统大赛提高了全球知名度

■ 网易云音乐

- 歌单推荐与评论推荐成为招牌

□ 推荐系统中包含的角色（核心三个）

人/用户

→ 消费者/内容贡献者：推荐对象

→ 有时人也会成为被推荐对象

物/商品

被推荐对象

非常广泛的概念

反馈/评价

→ 用户对于商品的评价，核心要素之一

→ 推荐的核心依据

→ 展现形式，两种

1) 显式反馈 (explicit feedback) : 评分; 评论; 标签; 点赞; 收藏等

2) 隐式反馈(implicit feedback): 购买次数; 页面停留时间; 眼球集中时间等

推荐系统中包含的角色

若干种反馈展现形式

- 评分：个人评分，总评分，分项评分
- 评论，可长可短；新形式：文字 + 图片
- 其它的类型比较常见，在此不再举例

速度与激情8 The Fate of the Furious (2017)



更新描述或海报

导演: F·加里·格雷

编剧: 克里斯·摩根 / 盖瑞·斯科特·汤普森

主演: 范·迪塞尔 / 道恩·强森 / 查理兹·塞隆 / 杰森·斯坦森 / 米歇尔·罗德里格兹 / 更多...

类型: 动作 / 犯罪

制片国家/地区: 美国

语言: 英语

上映日期: 2017-04-14(美国/中国大陆) / 2017-04-12(法国)

片长: 136分钟

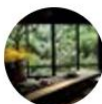
又名: 狂野时速8(港) / 玩命关头8(台) / 速激8 / 速8 / 速度8 / The Fast and the Furious 8 / Fast & Furious 8 / Fast Eight / Furious Eight / Furious 8 / Fast 8

IMDb链接: [tt4630562](https://www.imdb.com/title/tt4630562)

想看

看过

评价: ★★★★★ 力荐



临海沐月 🍷🍷🍷🍷🍷 VIP



口味: 4 环境: 4 服务: 4

前一阵吃的了，总体还是很满意的。餐厅整体设计很有自己的想法，服务也很棒。提前告知过生日什么的，会有布置。

【香煎鹅肝】鹅肝非常软腻，闪着着金黄色的鹅油，如.....

喜欢的菜：香煎鹅肝 银鳕鱼 当日限定

更多 ▾



查看全部9张

□ 推荐系统要素

■ 核心要素

- 人/用户、物/商品、反馈/评价

■ 非核心要素：!= 不重要要素

- 关联、内容、上下文

■ 推荐系统 != 某一类方法 或 单纯的理论/实验研究

- 推荐系统是一个完整的系统，是一类信息检索引擎
- 推荐系统包含多个组成部分，故我们会涉及很多非方法的部分
- 非方法的部分 != 单纯的背景，他们是推荐系统的组成部分

□ 推荐系统中包含的角色

关联

- 人与人之间的网络，可分为两种：直接关联与间接关联
- 物与物之间的关联
- 直接关联：
 - 1) 关注（单向，豆瓣，微博）
 - 2) 朋友（双向，人人，QQ空间）
- 间接关联：
 - 1) 同属于一组/话题（豆瓣，贴吧）
 - 2) 同购买一物品
 - 3) 被同一个人购买

内容

- 人的内容：人的各项属性，年龄，工作，收入，喜好等；又叫做user profile
- 物的内容：物的各项属性，名称，介绍，产地，价格，参数等
如电影的名字，介绍，导演，主演，配演，编剧，监制等
再如笔记本电脑的型号，参数，品牌、价格等
- 在内容数据缺失时，**评论**也可以作为内容的补充来源

□ 推荐系统中包含的角色

上下文/情境/Context: 时间, 地点, 情绪等

时间

- 时间对应部分商品的流通情况
- 一天中的时间: 我们一般不在中午推荐油条+豆浆
- 一年中的时间: 衣服, 旅游, 水果
- 节日时间: 电影, 旅游

地点

- 通过现在的手机定位, 可实现实时就近推荐
- 餐馆、电影院、出租车等就近推荐

情绪

- 主要针对交互式的物品, 比如音乐
- 高兴的时候就不要推荐感伤的
- 愤怒的时候可以推荐平和的

常用推荐方法

→ 推荐任务描述



西安电子科技大学
XIDIAN UNIVERSITY

□ 推荐任务描述

- 依据用户反馈或内容，从（大量）候选物品中，为用户预测出该用户可能交易/喜欢的物品
- 依据：用户反馈，主要表现形式包括 →
 - **显式反馈**：用户购买/收藏/转发/收看/收听记录与之后的反馈，一般为评分或者点赞
 - **隐式反馈**：用户购买次数，浏览记录，停留时间等
- 一般为大量物品
 - 推荐单位：可以为“个”，也可为“组”（歌单）
- 用户可能交易/喜欢
 - 推荐系统预测可能喜欢的物品，即为学习用户的兴趣

常用推荐方法

→ 推荐任务描述



□ 隐式反馈的重要性

- 显式反馈 (explicit feedback) 较为常见, 隐式反馈 (implicit) 在普通用户方面关注度不大
- 隐式反馈并非不重要, 甚至在一些场景中比显式反馈更重要
- 原因:
 - 一部分显式反馈不可靠
 - 评分、评论、点赞等存在的雇人刷分现象: 水军
 - 一部分评分、评论的给出并非出于兴趣, 而出于流程
 - 隐式反馈
 - 不容易出现作弊现象

■ 常用的推荐方法

- 基于协同过滤的推荐方法
- 基于内容的推荐方法
- 基于标签的推荐方法
- 基于网络的推荐方法
- 基于上下文的推荐方法

速度与激情8 The Fate of the Furious (2017)



更新描述或海报

导演: F·加里·格雷

编剧: 克里斯·摩根 / 盖瑞·斯科特·汤普森

主演: 范·迪塞尔 / 道恩·强森 / 查理兹·塞隆 / 杰森·斯坦森 / 米歇尔·罗德里格兹 / 更多...

类型: 动作 / 犯罪

制片国家/地区: 美国

语言: 英语

上映日期: 2017-04-14(美国/中国大陆) / 2017-04-12(法国)

片长: 136分钟

又名: 狂野时速8(港) / 玩命关头8(台) / 速激8 / 速8 / 速度8 / The Fast and the Furious 8 / Fast & Furious 8 / Fast Eight / Furious Eight / Furious 8 / Fast 8

IMDb链接: [tt4630562](https://www.imdb.com/title/tt4630562)

豆瓣评分

7.2  106613人评价

5星  16.0%
4星  38.9%
3星  37.0%
2星  6.5%
1星  1.6%

好友评分 

8.0  2人评价

好于 68% 动作片

好于 56% 犯罪片

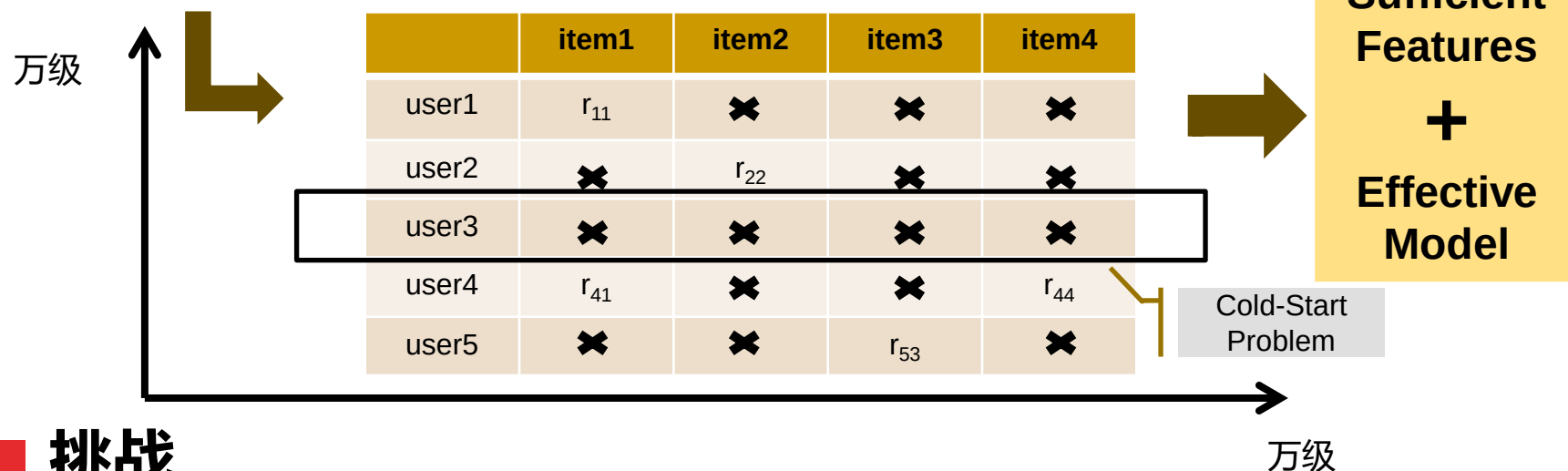
→ 基于协同过滤的推荐方法

□ 问题形式化

- Personalized rating prediction
- User-Item Rating Matrix/评分矩阵

两类典型问题：

- 评分预测问题
- Top-K推荐问题



挑战

- 数据稀疏性 → Cold Start Problem
- 数据规模巨大 → Feasibility

常用推荐方法

→ 基于协同过滤的推荐方法



□ 基于邻域的协同过滤推荐 (Neighborhood-based CF)

- 基于用户邻域: User-based CF algorithm
- 基于物品邻域: Item-based CF algorithm

□ 基于用户邻域的协同过滤:

	Shrek	Snow-white	Superman
Alice	Like	Like	Dislike
Bob	?	Dislike	Like
Chris	Like	Like	Dislike
Jon	Like	Like	?

□ Shall we recommend *Superman* to Jon?

- Jon的喜好与Chris、Alice都相似 → Do not recommend *Superman* to Jon

常用推荐方法

→ 基于协同过滤的推荐方法



□ 基于物品邻域的协同过滤推荐

■ Item-based CF algorithm

	Shrek	Snow-white	Superman
Alice	Like	Like	Dislike
Bob	?	Dislike	Like
Chris	Like	Like	Dislike
Jon	Like	Like	?

□ 推荐系统的对称性特点

常用推荐方法

→ 基于协同过滤的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 协同过滤方法的动机

- 靠其他人/物的力量，过滤掉不喜欢的物品/不被喜欢的用户
- 观察：For similar items or users:

- ✓ Users with dissimilar historical rating records usually have different preferences
- ✓ Users with similar historical rating records usually have similar preferences

- ✓ Items consumed by unsimilar users are usually received different ratings
- ✓ Items consumed by similar users are usually received similar ratings

■ 协同过滤的假设

- 历史记录中给出相似购买反馈的用户，在未来也可能具有相似的购买反馈
- 历史记录中具有相似的被购买反馈的商品，在未来也可能具有相似的被购买反馈

常用推荐方法

→ 基于协同过滤的推荐方法



□ 明确的目标：通过计算找到相似的用户或者物品

■ Explore similar users and items through historical ratings

■ 我们有什么？

➤ 我们有的就是这些historical ratings

➤ 评分矩阵/反馈矩阵，虽然很稀疏

	item1	item2	item3	item4
user1	r_{11}	?	?	?
user2	?	r_{22}	?	?
user3	?	?	?	?
user4	r_{41}	?	?	r_{44}
user5	?	?	r_{53}	?

常用推荐方法

→ 基于协同过滤的推荐方法



■ 第一步：如何计算相似性呢？（以皮尔逊相关系数为例）

➤ Pearson Correlation Coefficient → Similarity Computation

$$Sim(a, u) = \frac{\sum_{i \in I} (q_{ai} - \bar{q}_a)(q_{ui} - \bar{q}_u)}{\sqrt{\sum_{i \in I} (q_{ai} - \bar{q}_a)^2} \sqrt{\sum_{i \in I} (q_{ui} - \bar{q}_u)^2}}$$

users a 、 u 的相似性

$$Sim(i, j) = \frac{\sum_{u \in U} (q_{ui} - \bar{q}_i)(q_{uj} - \bar{q}_j)}{\sqrt{\sum_{u \in U} (q_{ui} - \bar{q}_i)^2} \sqrt{\sum_{u \in U} (q_{uj} - \bar{q}_j)^2}}$$

item i 、 j 的相似性

■ 第二步：相似邻居选择，构造相似邻居集

➤ 策略一：Top-K

➤ 策略二：Threshold

常用推荐方法

→ 基于协同过滤的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

■ 第三步：Predicted Missing Rating

➤ Self Average + Weighted Average (也有简单一些的计算方法)

$$p_{ui} = \bar{u} + \frac{\sum_{u_a \in S(u)} \text{Sim}(u, u_a)(r_{u_a, i} - \bar{u}_a)}{\sum_{u_a \in S(u)} \text{Sim}(u, u_a)}$$

基于用户邻域的协同过滤
User-based CF

$$p_{ui} = \bar{i} + \frac{\sum_{i_k \in S(i)} \text{Sim}(i, i_k)(r_{u, i_k} - \bar{i}_k)}{\sum_{i_k \in S(i)} \text{Sim}(i, i_k)}$$

基于物品邻域的协同过滤
Item-based CF

□ 总结

- 以上协同过滤方法称之为基于邻域的协同过滤
- Memory-based CF或者Neighborhood-based CF
- 下面进入model-based CF，即基于模型的协同过滤方法

常用推荐方法

→ 基于协同过滤的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 理想中的方法：奇异值分解

■ Singular Value Decomposition, SVD

■ For an arbitrary matrix $\mathbf{A}$, there exists a factorization SVD,

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}' \in \mathbb{R}^{n \times m}$$

$\mathbf{U} \in \mathbb{R}^{n \times k}$ $\mathbf{\Sigma} \in \mathbb{R}^{k \times k}$ $\mathbf{V} \in \mathbb{R}^{m \times k}$

$\mathbf{U}'\mathbf{U} = \mathbf{I}$ $\mathbf{V}'\mathbf{V} = \mathbf{I}$ *orthonormal columns*

$\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_k), \sigma_i \geq \sigma_{i+1}$ *singular values (ordered)*

$k = \text{rank}(\mathbf{A})$

常用推荐方法

→ 基于协同过滤的推荐方法

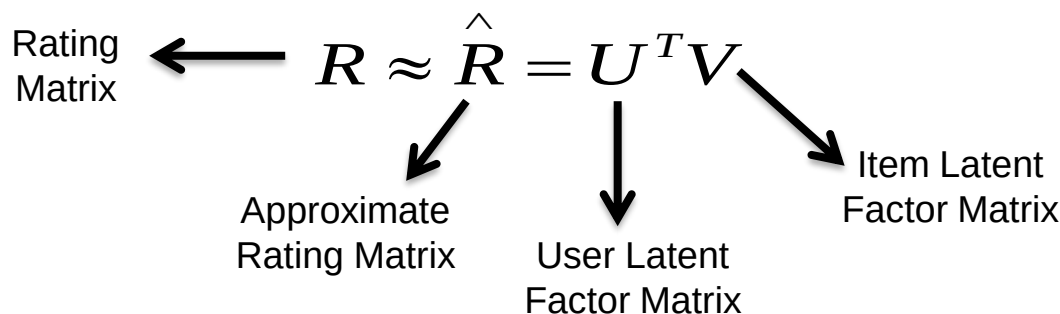


□ 实际的方法：隐因子模型（Latent Factor Model）

■ Latent Factor Space

□ 动机：You prefer *Lost in Thailand* because it' s a drama, and because X , and because Y , and because Z , and

■ X, Y & Z are named as **latent factors**



➤ Predicted value $\hat{R}(i, j)$

$$\hat{R}(i, j) = r(i, j) = \sum_k U_{uk} V_{ik} = \sum_k U(u, k) V(i, k) \quad \text{k-rank factors}$$

常用推荐方法

→ 基于协同过滤的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

最具有代表性的LFM → 矩阵分解
模型（家族）

常用推荐方法

→ 基于协同过滤的推荐方法



□ 矩阵分解: Matrix Factorization /Low-rank matrix factorization

- MF can factorize the high-rank user-item rating feature space into a joint low-rank feature space

- Basic MF模型

- $R=\{r_{ij}\}$: $m \times n$ user-item rating matrix

- $U \in \mathbf{R}^{d \times m}$, $V \in \mathbf{R}^{d \times n}$: user and item latent feature matrices

$r_{ij} \approx U_i^T V_j$ Inner product of two d -rank feature vectors

$$\min_{U,V} L = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (r_{ij} - U_i^T V_j)^2$$

Minimization function 1

I_{ij} , Indicator function(0/1)

常用推荐方法

→ 基于协同过滤的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

$$\min_{U,V} L = \underbrace{\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (r_{ij} - U_i^T V_j)^2}_{\text{Objective Function}} + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2$$

Regularization Term

Objective Function

□ Partial derivative and gradient descent algorithm

$$\frac{\partial L}{\partial U_i} = \sum_{j=1}^n I_{ij}^R (U_i^T V_j - r_{ij}) V_j + \lambda_u U_i$$

$$\frac{\partial L}{\partial V_j} = \sum_{i=1}^m I_{ij}^R (U_i^T V_j - r_{ij}) U_i + \lambda_v V_j$$

$$U_i = U_i - \eta \frac{\partial L}{\partial U_i}$$

$$V_j = V_j - \eta \frac{\partial L}{\partial V_j}$$

数据集较大时，也可以stochastic gradient descent

常用推荐方法

→ 基于协同过滤的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ Matrix Decomposition

- Tri-angle
- LU
- QR
- Spectral
- SVD

Matrix
Theory

■ Matrix Factorization

- Basic MF
- Non-negative
- PMF
- BPMF
- pLSA, LDA

Machine
Learning

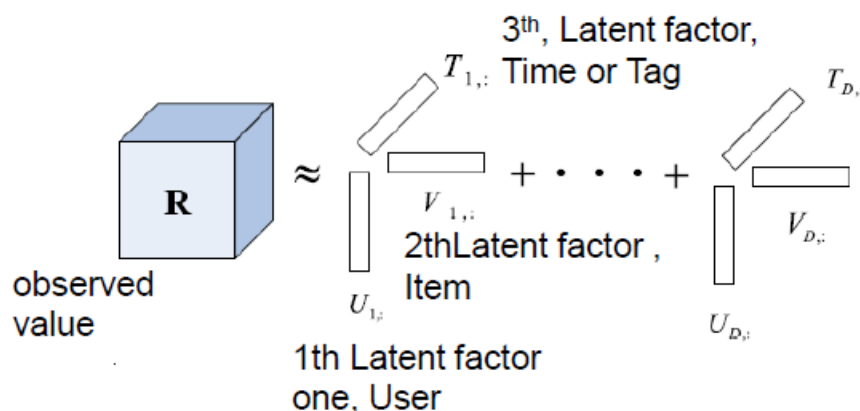
常用推荐方法

→ 基于协同过滤的推荐方法



■ 3-4-5- ∞ Dimensions → Tensor

- A tensor can be represented as a multi-dimensional array of numerical values.
 - 1-dimensional tensor : **Vector**
 - 2-dimensional tensor : **Matrix**
- Tensor Decomposition & ***Tensor Factorization***



常用推荐方法

→ 基于内容的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

推荐系统中的“冷启动”问题

□ 用户内容分析举例

■ 以iMDB上的电影/电视剧评分为例

■ iMDB：美国版的豆瓣

→ 男性与女性对于剧的喜爱有所不同（也有重叠）：

- 1) 喜爱类型有所不同
- 2) 喜爱程度有所不同

TOP SHOWS FOR MEN		TOP SHOWS FOR WOMEN	
SHOW	MALE RATING	SHOW	FEMALE RATING
Game of Thrones	9.5	Game of Thrones	9.4
Breaking Bad	9.5	Sherlock	9.4
The Wire	9.4	Breaking Bad	9.3
Rick and Morty	9.3	Last Week Tonight	9.3
The Sopranos	9.3	Avatar: The Last Airbender	9.3
Sherlock	9.2	Friends	9.2
Last Week Tonight	9.1	Firefly	9.2
Avatar: The Last Airbender	9.1	Rick and Morty	9.1
Fargo	9.1	Doctor Who	9.1
Firefly	9.1	The Wire	9.0
True Detective	9.1	True Detective	9.0
The Twilight Zone	9.0	The Twilight Zone	9.0
Monty Python's Flying Circus	9.0	Monty Python's Flying Circus	9.0
Arrested Development	9.0	Arrested Development	9.0
Batman: The Animated Series	9.0	Freaks and Geeks	9.0
Seinfeld	9.0	Gravity Falls	9.0
Only Fools and Horses	9.0	QI	9.0
House of Cards	9.0	Downton Abbey	9.0

常用推荐方法

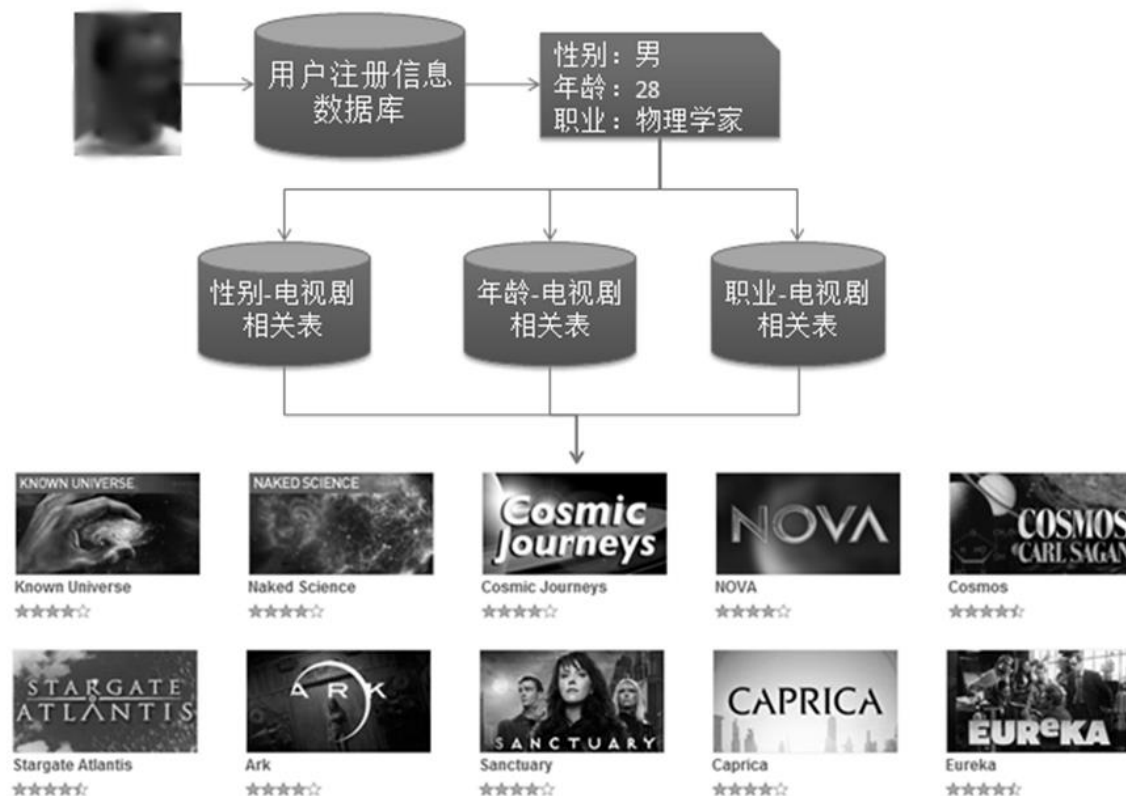
→ 基于内容的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 用户内容分析举例（续）

■ 职业影响（来自《推荐系统实践》）



常用推荐方法

→ 基于内容的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 物品内容

■ 举例

- 书籍：标题，作者，目录，正文
- 电影：名称，导演，演员，类别，简介
- 新闻：标题，作者/记者，正文，出处
- 电子产品：品牌，型号，配置，价格

■ 均可转化为文本

- 进一步转化为文本向量

题目	类型	作者	包装	价格	关键字
The Night of the Gun	Memoir	David Carr	Paperback	29.90	Press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism

常用推荐方法

→ 基于内容的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 基于用户内容与物品内容的计算

- 与物品内容相似，用户内容也可转化为文本向量

- 徐老师在豆瓣上的profile = (教师, 历史, 动作, 英雄, 战争)

- 徐老师在淘宝上的profile = (教师, 中产, 电子, 电脑, 日常)

- 计算用户内容与物品内容的相似程度

- 距离相似性, 空间相似性, 集合元素相似性

□ 推荐系统中的“冷启动”问题

- 用户：新用户或没有历史记录；物品：新物品或没有历史记录

- 基于内容的推荐特别适用于冷启动问题

常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 标签

- 无结构、用来描述物品的关键词

- 来源

- 物品提供方（作者/专家）：Pandora的Music Genome Project
- 普通用户：豆瓣，Delicious，Last.fm，Hulu，CiteULike
- 从评论中提取：京东，大众点评

- 物品提供方

Pandora的Music Genome Project标注出了400多个与歌曲有关的标签



常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

■ 普通用户提供

➤ 豆瓣



➤ 微博



常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

■ 从评论中提取

➤ 大众点评



➤ 京东



常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 标签的作用（来自 Shilads Wieland Sen 的博士论文）

- 表达：标签帮助表达对物品的看法（30%的用户认为）
- 组织：标签帮助组织喜欢的电影（23%的用户认为）
- 学习：标签帮助增加对电影的了解（27%的用户认为）
- 发现：标签使用户更容易发现喜欢的电影（19%的用户认为）
- 决策：标签帮助判定是否看某一部电影（14%的用户认为）

■ 调查平台：MoveLens

常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 标签的作用（来自 Shilads Wieland Sen 的博士论文）

- 表达：标签帮助表达对物品的看法（30%的用户认为）
- 组织：标签帮助组织喜欢的电影（23%的用户认为）
- 学习：标签帮助增加对电影的了解（27%的用户认为）
- 发现：标签使用户更容易发现喜欢的电影（19%的用户认为）
- 决策：标签帮助判定是否看某一部电影（14%的用户认为）

■ 调查平台：MoveLens →

常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 标签的作用（来自 Shilads Wieland Sen 的博士论文）

- 表达：标签帮助表达对物品
- 组织：标签帮助组织喜欢的
- 学习：标签帮助增加对电影
- 发现：标签使用户更容易发
- 决策：标签帮助判定是否看

■ 调查平台：MoveLens →

The screenshot displays the MovieLens interface for the movie 'The Shawshank Redemption'. The top section features a movie poster with a play button overlay. To the right of the poster, the movie title 'The Shawshank Redemption' is shown with a 4.29 star rating and a 'MovieLens predicts for you' section. Below the poster, a brief synopsis is provided: 'Framed in the 1940s for the double murder of his wife and her lover, upstanding banker Andy Dufresne begins a new life at the Shawshank prison, where he puts his accounting skills to work for an amoral warden. During his long stretch in prison, Dufresne comes to be admired by the other inmates -- including an older prisoner named Red -- for his integrity and unquenchable sense of hope.' The right sidebar contains metadata: '1994', '142 minutes', 'English', 'Frank Darabont' as director, and a cast list including Tim Robbins, Morgan Freeman, Bob Gunton, Clancy Brown, and Mark Rolston. It also lists the DVD release date as December 21, 1999, and provides links for movie maintenance. The bottom section, 'Your Tags', includes an 'add a tag' input field. The 'Community Tags' section shows a grid of tags with their counts, such as 'Morgan Freeman' (22), 'prison' (15), 'prison escape' (18), 'inspirational' (14), 'twist ending' (14), 'thought-provoking' (11), 'friendship' (13), 'classic' (11), 'Stephen King' (11), 'great ending' (9), 'drama' (8), 'hope' (7), 'justice' (7), 'great acting' (6), 'reflective' (7), 'atmospheric' (6), 'powerful ending' (4), 'based on a book' (6), 'heartwarming' (4), 'Tim Robbins' (5), 'imdb top 250' (4), 'revenge' (4), 'narrated' (4), and 'psychology' (3). The interface also includes a 'view' dropdown set to 'top' and a 'sort by' dropdown set to 'relevance'.

常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 标签清理与过滤

- 并不是所有的标签都是有用的

- 电影标签：“还可以”；音乐标签：“不好听”

- 标签格式需要规整

- Hip-hop, hip hop, Hip-Hop, hip_hop

□ 相似性比较方法

Step1: 统计每个用户最常用的标签

Step 2: 对于每个标签，统计被打过这个标签次数最多的物品

Step 3: 对于一个用户，首先找到他常用的标签，然后找到具有这些标签的最热门物品推荐给这个用户

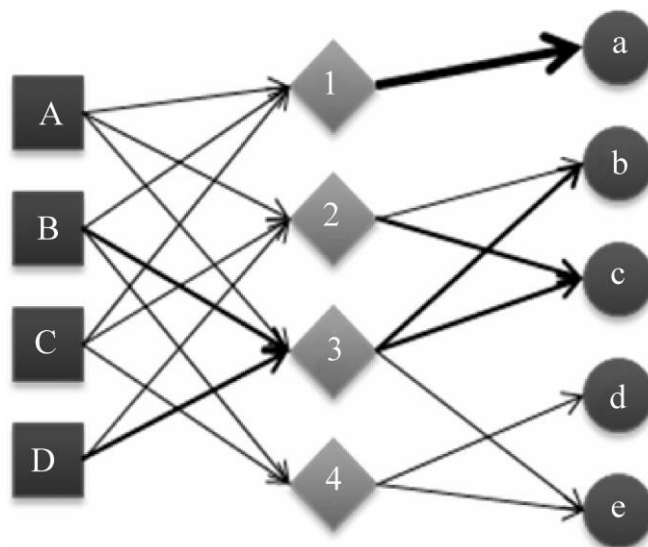
常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 进一步：基于图的标签推荐方法



- 可利用图的知识与算法，进行遍历、迭代与矩阵操作

常用推荐方法

→ 基于标签的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 延伸：标签推荐

- 为用户推荐可选可用的标签

速度与激情8 The Fate of the Furious (2017)

导演: F·加里·格雷
编剧: 克里斯·摩根
主演: 范·迪塞尔 / 米歇尔·罗德里格兹 / 杰森·斯坦森 / 卢克·伊万斯 / 道恩·强森 / 保罗·沃克 / 泰瑞斯·帕布森 / 卢克·布雷西 / 库尔特·拉塞尔 / 约翰·李 / 安东尼奥·班德拉斯 / 史蒂芬·加布里埃尔 / 斯科特·凯恩 / 詹姆斯·布兰登 / 布莱恩·泰 / 小尼尔·布朗 / 阿比盖尔·斯宾赛 / 塔拉·李 / 乔纳森·海德 / 大卫·斯塔马尔茨 / 萨姆·鲁宾 / 布莱恩·布鲁姆 / 斯科特·迈克尔·坎贝尔 / 贾斯汀·巴萨 / 杰克·布塞 / 凯瑟琳·温妮克 / 乔什·杜哈明 / 安东尼·迈克尔·弗雷德里克 / 内森·克拉克 / 尼克·拉加 / 安德鲁·康斯坦丁尼 / 史蒂夫·皮特森 / 马特·史密斯 / 艾丽卡·克里奇 / 贾斯汀·朗 / 马修·卢卡斯 / 大卫·安德鲁 / 丹·皮特森 / 大卫·斯特雷泽 / 瑞安·阿罗索 / 博·米尔博夫 / 史蒂夫·库珀 / 杰森·乔治 / 凯文·史密斯 / 大卫·惠特 / 凯特·伯顿 / 戴安娜·加德纳 / 维克多·加博 / 迈尔斯·杰弗里 / 杰夫·柯蒂斯 / 艾丽西娅·维拉-贝利 / 詹姆斯·肖安科 / 阿隆·拉扎里 / 比利·克鲁德普 / 杰西卡·科林斯 / 格雷厄姆·J·福特 / 迈克尔·杰特 / 琳恩·柯林斯 / 阿德里安·迪安 / 杰夫·库普伍德 / 维托·达布罗西 / 弗兰克·迪安 / 米凯莱·普斯特里尼 / 杰夫·伯格 / 阿曼达·诺顿 / 阿伦·戴维森 / 斯蒂芬妮·索科琳斯基 / 珍妮·卡斯 / 布赖恩·克卢格曼 / 凯蒂·加洛 / 乔纳森·格雷格 / 琼·瓦勒拉 / 马克·瑞安 / 阿德里安·鲁特 / 杰瑞米·霍华德 / 蒂姆·霍尔姆斯 / 布兰登·斯平克 / 贾斯汀·朗 / 詹姆斯·华纳 / 罗宾·哈特 / 杰西·加西亚 / 特雷尔·蒂尔福德 / 丽莎·奥图尔 / 史蒂文·韦斯 / 大卫·康拉德 / 约翰·派珀-弗格森 / 斯科特·劳伦斯 / 梅丽莎·西尔维奥 / 斯科特·威廉姆·文特斯 / 汉娜·韦斯特 / 詹姆斯·汉内 / 乔·阿西姆 / 米歇尔·莫纳汉 / 小欧内斯特·史密斯 / 伊桑·苏普利 / 扎克·杜拉磨 / 艾米莉·科迪 / 崔·金 / 梅根·查彭迪厄 / 迪伦·阿诺德 / 莱昂·里奇 / 詹姆斯·肖安科 / 帕特里卡·达博 / 凯特·沃什 / 阿梅莉娅·沃森 / 梅尔·埃斯皮诺萨 / 安妮·凯利 / 梅丽莎·乔伊斯 / 杰夫·沃德 / 萨米·谢伊克 / 乔什·杜哈明 / 安东尼·麦凯 / 阿曼达·舒尔 / 约翰·黑泽尔 / 乔纳森·亨利 / 乔安娜·布拉迪 / 斯蒂芬·托布罗斯基 / 科迪·沙尔 / 梅罗妮·迪亚兹 / 瑞安·艾戈尔德 / 布兰登·司徒 / 尼克·阿拉波格鲁 / 杰克·麦克道曼 / 科迪·林恩 / 乔·曼根尼罗 / 约翰·奥提兹 / 布鲁斯·麦克吉尔 / 克里斯·加恩 / 斯科特·麦克纳尼 / 雷吉·李 / 安吉拉·麦克曼纳 / 蒂娜·霍尔姆斯 / 安德鲁·李 / 杰里·哈德森 / 马克·伊瓦涅 / 克里斯托弗·米洛尼 / 迪伦·麦克德莫特 / 厄尼·里夫利 / 布莱恩·利利斯 / 达斯蒂·马尔蒂斯 / 瑞克·范维斯 / 斯蒂夫·安德森 / 亚当·利维 / 斯凯勒·迪恩 / 埃里克·安德烈 / 约翰·巴罗曼 / 内森·达格 / 迪恩·J·韦斯特 / 乔·曼根尼罗 / 乔纳森·利普斯科特 / 乔恩·沃茨 / 克里斯·康纳 / 凯尔·塞缪尔逊 / 本·阿勒斯 / 罗德里戈·桑托罗 / 妮可·戈登 / 罗斯·布莱克维尔 / 露西·麦卡锡 / 丹尼尔·伯恩斯 / 布莱恩·格林 / 詹姆斯·兰德里·赫伯特 / 阿比盖尔·菲茨杰拉德 / 埃文·埃瓦戈拉 / 特雷西·邦纳 / 艾米·菲灵顿 / 雪拉·渥德 / 大卫·芬奇 / 乔·曼根尼罗 / 杰伊·巴鲁切尔 / 伊尔凡·可汗 / 凯尔·麦克拉克伦 / 大卫·斯特雷泽 / 瑞安·艾戈尔德 / 杰西·威廉姆斯 / 达蒙·欧普瑞 / 詹姆斯·克伦威尔 / 艾利克斯·迪塞尔 / 丹尼尔·罗德里格斯 / 诺亚·泰勒 / 卡梅隆·米切尔 / 凯瑟琳·波切尔 / 迈克·沃格尔 / 凯文·麦克纳尔蒂 / 蒂莫西·布斯菲尔德 / 史蒂夫·巴斯默 / 乔什·杜哈明 / 安东尼·麦凯 / 克莱尔·布朗 / 伊莎贝拉·莫塞德 / 马特·瑞安 / 卡梅伦·克拉克 / 杰夫·道切特 / 乔什·塞巴斯蒂安 / 崔西·摩根 / 詹姆斯·甘多菲尼 / 罗伯特·约翰·伯克 / 斯科特·波特 / 盖瑞·科尔 / 托尼·丹茨 / 艾米莉·方安 / 杰西卡·皮科拉 / 杰弗里·布莱克 / 艾莉森·托尔曼 / 梅雷迪思·加勒特 / 康纳·莱斯利 / 罗伯特·帕特里克 / 亚当·佩里 / 维吉尼亚·威廉姆斯 / 考特尼·卡尔弗特 / 米洛·文堤米利亚 / 梅兰妮·林斯基 / 阿萨夫·凯斯 / 伊莱恩·贾夫瑞 / 凯瑟琳·纳杜奇 / 丹·皮克 / 詹姆斯·罗宾逊 / 史蒂夫·佩梅 / 杰西卡·查斯坦 / 伊莎贝拉·阿西拉 / 萨米拉·威利 / 迪耶戈·博内塔 / 卡门·阿尔扎雷戈 / 雷欧·蒂瓦兹 / 伊娃·阿里山丝 / 马克·印第里凯托 / 阿曼达·赛克斯 / 索非亚·赫芬顿 / 马特·瑞安 / 杰夫·桑吉斯 / 艾丽莎·坎伯 / 弗兰克·威利 / 琳恩·柯林斯 / 约瑟夫·沃克 / 大卫·圣托勒 / 迈克尔·肖沃特 / 凯蒂·帕克 / 斯科特·德克特 / 李·佩斯 / 贝蒂·怀特 / 阿黛莱德·凯恩 / 伊莎贝拉·布莱克 / 安妮·帕里西 / 布鲁斯·麦克吉尔 / 罗德里戈·桑托罗 / 汤姆·维里卡 / 弗兰克·格里罗 / 詹姆斯·沃克 / 布莱恩·泰 / 乔什·塞巴斯蒂安 / 约瑟夫·坎农 / 马克·康苏斯 / 布莱恩·马克金森 / 伊莱恩·帕里 / 杰西卡·刘 / 邦尼·阿伦斯 / 阿隆·希尔 / 迈尔斯·巴克利 / 比利·斯洛特 / 德鲁·鲍威尔 / 杰伊·布莱恩·史蒂文森 / 安妮·帕里西 / 朱迪斯·希格德 / 索非亚·伯尔曼 / 丽莎·坎宁 / 斯科特·德克特 / 塞巴斯蒂安·阿梅斯托 / 亚当·佩里 / 凯蒂·洛茨 / 格伦妮·海德利 / 丽贝卡·理查兹 / 汤姆·佩德罗兹 / 斯科特·劳伦斯 / 布鲁斯·邓恩 / 凯瑟琳·威利斯 / 萨赫丽玛 / 科雷恩·雷诺兹 / 塞巴斯蒂安·安托里尼 / 苏芮琪 / 约翰·派珀-弗格森 / 菲丽西亚·戴 / 玛德琳·奇玛 / 希罗·弗南德兹 / 亚历克斯·怀特 / 路易斯·米格尔·阿尔卡拉 / 加布里埃尔·伯格曼 / 马克·沃尔什 / 梅乐迪丝·梦露 / 蒂凡尼·法比亚诺 / 塞德里克·亚伯勒 / 帕特里克·赫辛格 / 艾丽·范宁 / 塞巴斯蒂安·马尼斯科 / 夏奈尔·塞西尔 / 阿梅莉娅·沃森 / 詹姆斯·克伦威尔 / 米基·洛基 / 崔德·沃特雷 / 詹姆斯·麦凯 / 琳恩·柯林斯 / 乔恩·坎农 / 朱利安·巴恩斯 / 杰米·坎农 / 斯蒂夫·蒙罗埃 / 塞德里克·亚伯勒 / 阿伦·戴维森 / 马迪森·麦克洛林 / 阿曼达·福斯特 / 迈克·梅霍尔 / 卡勒姆·肖恩 / 塞西尔·威廉姆斯 / 塞缪尔·哈尔维尔 / 伊莎贝拉·阿西拉 / 凯特·玛拉 / 约翰·卡伦 / 琳迪·卡特 / 比利·斯洛特 / 乔什·诺顿 / 约翰·迪尔 / 杰瑞·诺兰 / 萨米·纳塞利 / 布鲁斯·麦克吉尔 / 阿德里安·鲍威尔 / 琪兰·席普卡 / 爱德华多·科斯塔 / 玛丽亚·巴卡洛娃 / 乔安妮·博兰德 / 布伦丹·凯利 / 史蒂芬·奥斯汀·莫里斯 / 加布里埃尔·鲁特 / 阿曼达·布鲁盖尔 / 杰瑞米·艾恩斯 / 克里斯托弗·阿波特 / 杰夫·科贝特 / 迈克尔·贝利·史密斯 / 凯文·杜恩 / 马修·詹姆斯·凡·皮特斯 / 海莉·博内特 / 朱丽安娜·古尔 / 杰西卡·亨维克 / 安妮·帕里西 / 李·佩斯 / 丹·佩罗 / 莉迪亚·波尔图 / 辛西娅·罗伯茨 / 凯瑟琳·普雷斯科特 / 卢克·马可法莱蒂 / 罗杰·巴特 / 托尼·丹兹 / 梅根·加拉格尔 / 杰米·坎农 / 安娜·迪奥普 / 马克斯·凯塞拉 / 戴安娜·斯卡尔维德 / 斯特凡妮娅·罗默 / 萨安娜·威尔逊 / 萨姆·斯威兹 / 艾格尼丝·布鲁根 / 阿梅莉娅·沃森 / 菲奥娜·格林 / 萨拉·坎宁 / 伊丽莎白·坎 / 尼克·维斯 / 詹姆斯·萨利 / 杰奎琳·罗伯茨 / 亚德里安·帕斯达 / 乔舒亚·拉塞姆 / 马克·达卡斯考斯 / 艾丹·吉伦 / 伊芙·普拉姆 / 戴斯敏·博格斯 / 乔丹·卢戈 / 卢克·马可法莱蒂 / 皮耶尔弗兰切斯科·法维诺 / 马修·詹姆斯·凡·皮特斯 / 马克·斯坦利 / 卢卡斯·索比诺 / 伊恩·莱恩 / 肖恩·麦克米伦 / 丹·莱瑟姆 / 加布里埃拉·布罗克 / 戴安娜·维利加斯 / 詹姆斯·乔恩 / 理查德·坎德 / 泽维尔·塞缪尔 / 斯图尔特·卡本尼 / 布鲁斯·巴克林纳 / 肖恩·多伊尔 / 杰瑞米·艾恩斯 / 凯文·麦克哈尔 / 安娜·克拉姆斯基 / 欧文·坎贝尔 / 布兰登·罗宾逊 / 约翰·奥利维 / 乔伊·科恩 / 瑞安·瑞特 / 库尔特·纽曼 / 维德尔·皮尔斯 / 盖瑞·科尔 / 阿西夫·阿里 / 威尔·法瑞尔 / 约翰·斯拉特里 / 欧文·威斯克 / 阿曼达·诺顿 / 萨拉·赖特 / 乔伊·科恩 / 提姆·里奇 / 詹姆斯·迈克尔·麦考蒂 / 黛比·瑞奇 / 希拉·凡德 / 肖恩·基尔西 / 查尔斯·C·史蒂文森 / 达里尔·斯蒂芬斯 / 艾萨克·麦迪逊 / 卡梅伦·克拉克 / 杰夫·普罗斯特 / 杰西·平托 / 卢克·柯比 / 莱斯利·格罗斯曼 / 朱莉·托尔 / 凯瑟琳·梅塞尔 / 玛丽·帕特·格里森 / 詹姆斯·迪伦 / 吉奥·瑞基 / 约翰·金 / 尼克·扎诺 / 科里·哈德里克 / 斯科特·德克特 / 马克·斯特朗 / 阿兰·卢克 / 雷恩·威尔森 / 杰姬·摩尔 / 詹姆斯·霍兰 / 阿德里安·斯卡迪诺 / 大卫·斯塔马尔茨 / 罗伯特·维斯多姆 / 迈克尔·切赫斯 / 杰伊·巴鲁切尔 / 比利·坎贝尔 / 阿曼达·皮特 / 斯科特·波特 / 阿兹·泰 / 加里·斯普拉蒙 / 卡洛琳·沃西尔 / 肯尼斯·以色列 / 阿尔比·海尔 / 卡米拉·基鲍 / 罗德·拉里·约翰逊 / 珍·斯马特 / 朗·罗格 / 瑞恩·多诺西 / 凯文·诺兰 / 迪恩·诺里斯 / 萨姆·乔希 / 维罗尼卡·卡维特 / 安迪·里克特 / 克里斯托弗·科辛斯基 / 达伦·皮尔斯 / 肖恩·布朗 / 伊丽莎白·班克斯 / 贾斯汀·沃德尔 / 查尔斯·哈尔福德 / 乔纳森·尼科尔斯 / 乔琳·珀迪 / 洛伦佐·詹姆斯·亨里 / 安娜莱吉·阿什福特 / 玛德琳·科莱特 / 达伦·肯特 / 詹姆斯·盖蒙 / 梅利莎·古德温 / 奥古斯特·理查兹 / 菲利普·卡斯诺夫 / 哥林·费舍尔 / 霍莉·维恩 / 琳恩·麦克阿瑟 / 崔德·沃特雷 / 乔伊·纳什 / 迈克尔·诺里 / 格蕾丝·拜尔斯 / 迈克·安东纳科斯 / 德米特里·斯库尔采夫 / 凯文·赫夫南 / 亚历克西斯·拉莫斯 / 杰夫·道切特 / 科尔·雷菲尔德 / 帕特里克·茂顿 / 杰夫·科贝特 / 伊万·奥图尔 / 戴安娜·劳伦 / 安妮·波莉 / 詹姆斯·瓦伦汀 / 詹姆斯·乔丹 / 达里尔·迪莱尼 / 杰西卡·苏拉 / 瓦妮莎·贝尔 / 斯科特·德克特 / 乔安娜·加西亚 / 乔什·德伦 / 阿德里安·克鲁格 / 丹·艾克罗伊德 / 加里·维克斯 / 麦克·瓦尔茨 / 杰夫·威勒 / 凯文·赫夫南 / 迈克·安东纳科斯 / 德米特里·米切尔 / 凯文·奥罗克 / 约翰·李 / 安娜玛丽亚·罗拉 / 拉马尔·约翰逊 / 戴维·J·维斯 / 约翰·休斯 / 斯科特·施密特 / 马克·卡西迪 / 阿曼达·塞弗里德 / 斯蒂芬·凯利 / 詹姆斯·T.沃德 / 美国 / 詹姆斯·曼高德 / 136 分钟 / 152 分钟(导演剪辑版) / 速度与激情8 / 动作 / 犯罪 / 迈克尔·格林 / 盖瑞·斯科特·汤普森 / 理查德·林德赛 / 克里斯托弗·麦考里 / 英语

■ 作用

- 1) 方便用户输入; 2) 提高标签质量

常用推荐方法

→ 基于网络的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 网络：在此指用户之间的关系

■ 社交网络（关注）

- 纯社交网络（双向/单向关系）：微博，Facebook，Twitter
- 社区网络（单向关系）：豆瓣，淘宝，大众点评

■ 信任网络：Epinions.com

- 我对你的打分信任，我对你的打分不信任
- 数据集贡献大户，单向关系

■ 隐含网络（双向关系）

- 共同在一个小组，共同写过一篇论文，共同

我的关注 ····· (成员32 · 小站15)



丕子



Alon



王逅逅



赵小凡



飞林沙



C6H5NO2



易水寒



豆瓣吃货

常用推荐方法

→ 基于网络的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 基本假设

- 用户的兴趣爱好与他/她朋友们（或部分）的兴趣爱好具有一定程度的相似性

■ 举例

- 你的朋友正在听的音乐（QQ音乐，网易云音乐）
- 你的朋友正在玩的小游戏（QQ游戏）
- 你关注的人发布或转发的微博，或者关注的动态（新浪微博，Twitter）
- 你所在小组成员关注的其它小组（豆瓣）
- 你所关注的达人所去的餐馆或商店（大众点评）

常用推荐方法

→ 基于网络的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 基本假设

- 用户的兴趣爱好与他/她朋友们（或部分）的兴趣爱好具有一定程度的相似性

■ 举例

- 你的朋友正在听的音乐（QQ音乐，网易云音乐）
- 你的朋友正在玩的小游戏（QQ游戏）
- 你关注的人发布或转发的微博

好友动态

部落动态

默认 点评



下一站的彩虹 点评了 绿茶

口味：4 环境：4 服务：4

今天天气不错，我和远道而来的好朋友逛完了美景，我们再一次坐在这里品尝美味佳肴，我和我的好朋友都非常喜欢绿茶的环境和美食，又点了她念念不忘的面包诱惑😋，好朋友说这边的菜价格实惠又好吃，服务还非常到 [全文]

04-26 13:15

这个小组的数字狂也喜欢去

 统计学 应用 问答: SPS S... (6085)	 数据挖掘 (14004)	 R (GNU S) (3335)	 数学 Mathe matics (35992)
 变态精算师 (4558)	 Financial En gineering (4645)	 纯数学 (12579)	 数据可视化 Data Visua... (1780)

好友关注动态

neK 等760万人关注了

SGT_SONG 关注
微博知名军事博...

+ 关注

KUCS 等4人也关注

全球旅游攻略 关注
知名旅游博主

+ 关注

常用推荐方法

→ 基于网络的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 好友重要性/依赖性

■ 作出推荐时，应该更依靠用户的哪些好友的喜好？

➤ 基本准则：相似度高的，但也不一定

■ 社交网络的另一种分类方法

➤ **熟人网络**：人人网，QQ空间，微信朋友圈等双向

➤ **陌生人网络**：微博、豆瓣关注、点评关注、淘宝的网络

■ 推荐策略区别

➤ 熟人网络：寻找“相似性”最高的好友

➤ 陌生人网络：寻找“权威性”最高的专家（大V）



他的主页

常用推荐方法

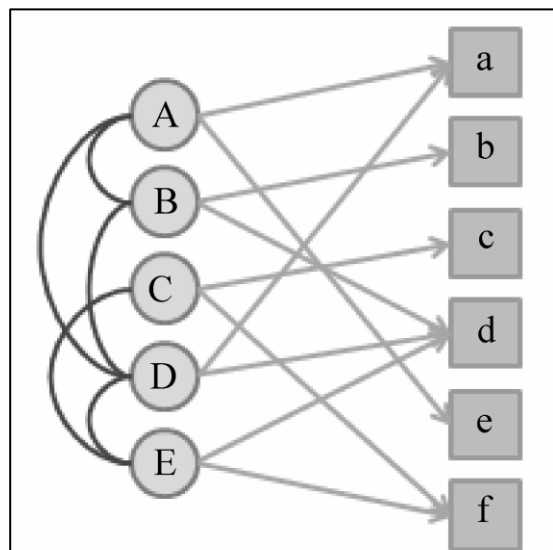
→ 基于网络的推荐方法

□ 基于社交网络的推荐方法

■ 也叫做社会化推荐

➤ 用户与用户之间的各种关系是社会化交互的线上反映

□ 基于图的社会化推荐方法



➔ 用户社交网络与用户-物品购买图

➔ 利用基于图的方法进行推荐，类似于基于标签的推荐方法

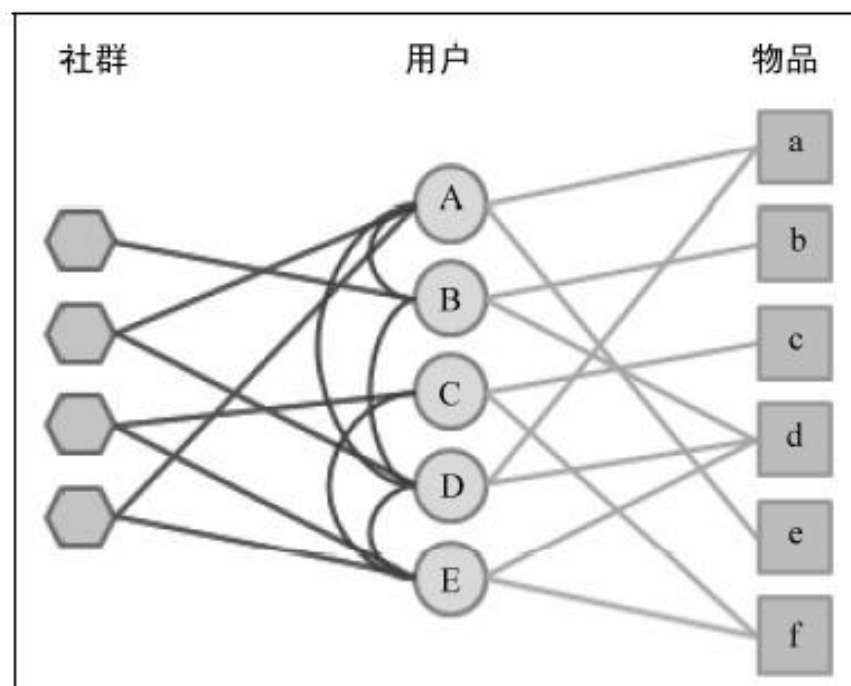
➔ 用户-用户关系矩阵与用户物品交互矩阵



常用推荐方法

→ 基于网络的推荐方法

□ 基于图的复杂社会化推荐方法



→ 用户与用户之间的线：
社交关系，一般为双向社交关系

→ 社群指代社区关注，用户为社区/小组中的成员

→ 社区/小组中的关系又叫做“成员关系”

→ 用户-用户关系矩阵，
用户-社区关系矩阵，用户物品交互矩阵

常用推荐方法

→ 基于网络的推荐方法



■ More Relationship → More Matrix

➤ Social Network

- Turn to your friends for suggestion

➤ Trust Network

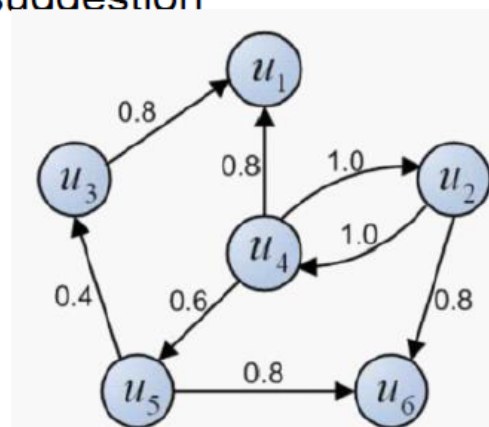
- Turn to who you trust for suggestion

➤ Clarify the connection

- What's the relationship?
- Why does it work?

■ Etc.

➤ Structure of Networks



**Weight &
Relationship**

**Social/Trust
Network**

常用推荐方法

→ 基于上下文的推荐方法



□ 什么是上下文 (Context)

■ 用户与推荐系统交互时的时间、地点、心情、人群等因素

- 即回答“用户是在什么情况下与推荐系统进行交互的”
- 上下文信息可能会呈现出不同的粒度

■ 举例

- 时间：一年中的季节，用户兴趣的变化，物品特征的变化
- 地点：在家，在商场，在影院；在海南，在哈尔滨，在阿尔泰
- 心情：很高兴，高兴，悲伤，很悲伤
- 人群：和父母一起看电视，和女朋友一起看电视，你自己一个人看电视

常用推荐方法

→ 基于上下文的推荐方法

□ 真实应用举例

- Sourcetone的音乐推荐系统会提供情绪选项
- 大众点评、美团等推荐附近的餐馆与电影院
- 麦当劳根据一天的时间推荐套餐
- 电影院根据人群的不同推出的专场



常用推荐方法

→ 基于上下文的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 时间上下文

■ 用户兴趣的变化

- 软件工程师：从当当上买的书从一开始的《XXX入门》到《XXX精通》到《XXX核心》
- 博士生：毕业之前买电脑侧重计算性能，毕业之后买电脑侧重生活性能

■ 物品生命周期

- 电影的上映期，畅销书的销售期，一首歌的流行期

■ 季节性与节日性

- 衣服的季节性、各个节日时的礼物（圣诞，双11），各种销售节（家装节，电脑节，开学日）

常用推荐方法

→ 基于上下文的推荐方法



西安电子科技大学
XIDIAN UNIVERSITY

□ 地点上下文

■ 重要的应用：基于位置的服务

- 推荐餐馆等：存在于大众点评，美团，饿了么，糯米，百度地图，Yelp等中的O2O功能中
- 推荐好友等：附近的人

■ 用户的兴趣/活动与地点也有一定的关系

- 兴趣本地化：西安的观众更喜欢西安本地出身的演员与导演（可能）；西安的老百姓更喜欢凉皮+辣子
陕南喜欢秦腔，陕北喜欢信天游
- 活动本地化：看电影一般都是选比较近的；节假日一般都是选远一点的



□ 学术界如何评测推荐方法

■ Evaluation measure

- RMSE
- MAE
- NMAE
- Precision
- Recall
- NDCG
- .etc

□ 工业界如何预判一种推荐方法是否起作用？

■ A/B测试（通用方法）

■ 随机（双盲）对照实验，举例：

- 随机将两个不同的内测版本发布给两组相同的，随机选择出的两组用户
- 两个版本的推荐系统不同：
 - 所用方法不同（有基于内容的，有基于网络的）
 - 有一个版本为旧的，一个版本为新的



□ 个人主页

■ 课件（包含推荐系统）

- <http://web.xidian.edu.cn/ysxu/teach.html>
- <http://www.slideshare.net/obamaxys2011>

□ 各推荐系统比赛

■ Netflix全球推荐系统比赛（2008）

- 极大地推动了推荐系统本身的关注度

■ Kaggle上由Yelp等公司发起的推荐系统比赛

■ 阿里天池每年举办的推荐系统比赛

各推荐系统比赛



西安电子科技大学
XIDIAN UNIVERSITY

□ Netflix推荐系统比赛



各推荐系统比赛



西安电子科技大学
XIDIAN UNIVERSITY

□ Kaggle

The screenshot shows the Kaggle website interface for the RecSys2013: Yelp Business Rating Prediction competition. The top navigation bar includes the Kaggle logo, a search bar, and links to Competitions, Datasets, Kernels, Discussion, Jobs, and a user profile icon. The competition title is "RecSys2013: Yelp Business Rating Prediction", associated with "The ACM Conference Series on Recommender Systems". It is a "RecSys Challenge 2013: Yelp business rating prediction" with a prize of "\$500" and "158 teams" participating "4 years ago". The page has tabs for Overview, Data, Discussion, Leaderboard, and More, with a "Submit Predictions" button. The Overview section is active, showing a sidebar with links to Description, Evaluation, Prizes, Timeline, and Winners. The main content area features a large image of a city skyline at sunset.

各推荐系统比赛



西安电子科技大学
XIDIAN UNIVERSITY

□ 阿里天池比赛

首页 > 天池大赛 > 阿里移动推荐算法



黄金联赛

状态

举办方

第 2 赛季截止日期

总奖池

参赛队

阿里移动推荐算法

已结束



2015/07/01

¥ 300000

7186

赛制介绍

赛题与数据

FAQ

排行榜

论坛

MobileDM 2015

赛题简介

2014年是阿里巴巴集团移动电商业务快速发展的一年，例如2014双11大促中移动端成交占比达到42.6%，超过240亿元。相比PC时代，移动端网络的访问是随时随地的，具有更丰富的场景数据，比如用户的位置信息、用户访问的时间规律等。

本次大赛以阿里巴巴移动电商平台的真实用户-商品行为数据为基础，同时提供移动时代特有的位置信息，而参赛队伍则需要通过大数据和算法构建面向移动电子商务的商品推荐模型。希望参赛队伍能够挖掘数据背后丰富的内涵，为移动用户在合适的时间、合适的地点精准推荐合适的内容。

赛制安排

第一赛季，3月20日—4月25日

1. 可下载少量淘宝数据，并在本地调试算法，提交结果；若队伍一天内多次提交结果，新结果版本将覆盖原版本；



- 计算广告学
 - 广告竞价
 - 广告推荐
- 商品/物品生命周期预测
- 网络中的专家发现
- 网络中的社区发现
- 评论观点挖掘
- 评论主题挖掘
- 虚假评论挖掘



西安电子科技大学
XIDIAN UNIVERSITY

Q&A