数据挖掘导论
Introduction to Data Mining

第六章 推荐系统

刘 淇
Email: qiliuql@ustc.edu.cn
课程主页：
http://staff.ustc.edu.cn/~qiliuql/DM2017YZ.html

主要目标

- 了解推荐系统的背景和现状
- 总结推荐算法的主要思想、技术方案
- 建立推荐系统与数据挖掘的关联
- 学习推荐算法的设计（研究）案例
- 探讨推荐算法的未来研究方向

主要内容

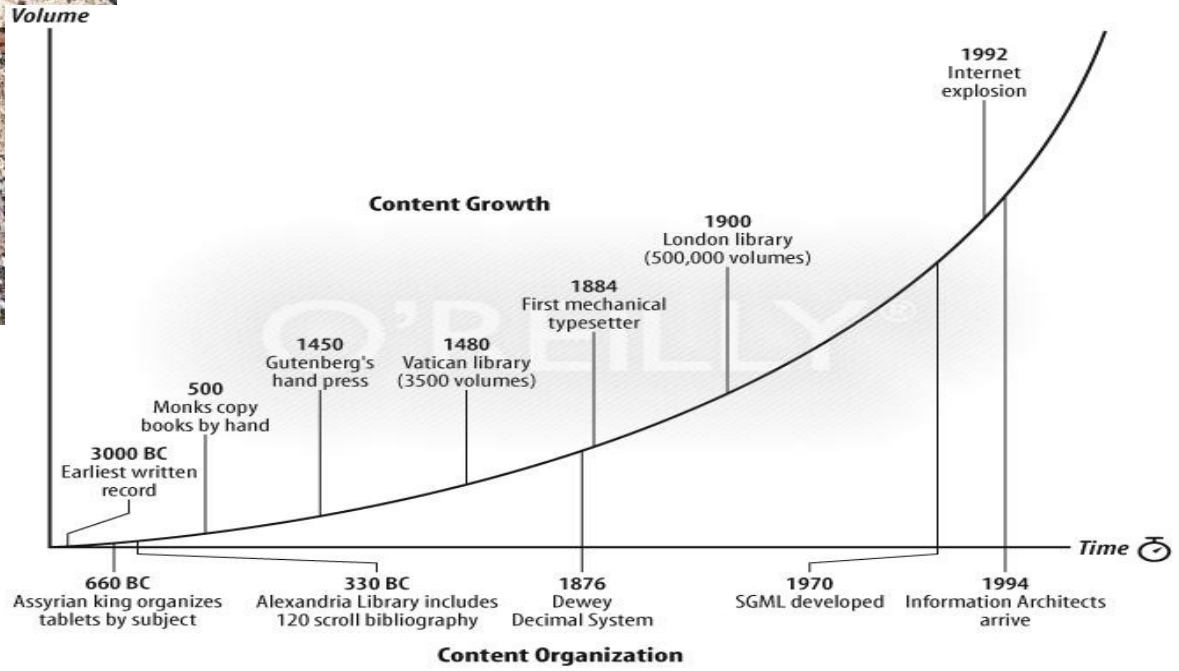
- 什么是推荐系统
  - 背景、定义、应用场景
- 推荐方法概述
  - 兴趣建模
  - 推荐算法设计
  - 推荐结果的评估
- 案例学习
  - 基于用户兴趣扩展的个性化推荐方法
  - 面向推荐系统的纠结心理挖掘
- 小结及未来的路
- 资料推荐

主要内容

- 什么是推荐系统
  - 背景、定义、应用场景
- 推荐方法概述
  - 兴趣建模
  - 推荐算法设计
  - 推荐结果的评估
- 案例学习
  - 基于用户兴趣扩展的个性化推荐方法
  - 面向推荐系统的纠结心理挖掘
- 小结及未来的路
- 资料推荐

为什么要推荐

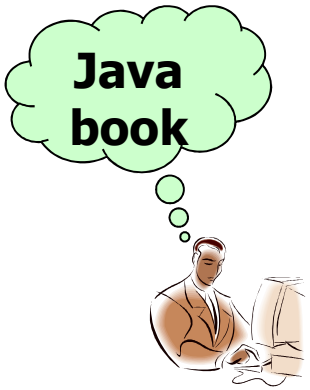
Information overload



We are leaving the age of information and entering the age of recommendation.
Chris Anderson in *The Long Tail*

背景

- 原因示例1

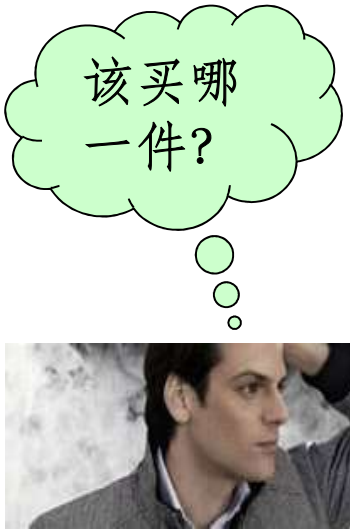


信息推荐！！

返回信息量太大

背景

- 原因示例-2



个性化信息推荐！！

背景

- 推荐系统

信息技术

网络商家

项目

普通用户

推荐系统

- 项目过滤
- 主动服务

商家增加收益

用户各得所需

推荐无处不在—豆瓣

I Am Legend

简体中文名: 我是传奇

编剧: Mark Protosevich / Akiva Goldsman / Richard Matheson
导演: Francis Lawrence
主演: Will Smith / Alice Braga / Charlie Tahan
官方网站: http://iamlegend.warnerbros.com/
上映年度: 2007
语言: 英语
制片国家/地区: 美国
imdb链接: tt0480249

放在你的blog里!

增改描述、海报图片

我看过这部电影 修改 删除

我的评价: ★★★★★力荐

★★★★★ 2851
★★★★☆ 8146
★★★☆☆ 6643
★★☆☆☆ 968
★☆☆☆☆ 140

推荐

豆瓣猜你可能感兴趣的电影

喜欢看"这部电影"的人也喜欢:

机械公敌

全民超人


国家宝藏2：古籍秘辛

通缉令


科洛弗档案

钢铁侠

心灵传输者


300


迷雾


国家公敌

300 / 300死士 / 300斯巴达勇士

Gerard Butler / Vincent Regan / Lena He

看过 想看 没兴趣

Iron Man / 铁人 / 钢铁侠

Robert Downey Jr. / Terrence Howard /
钢铁侠 / Art Marcum / Matt Holloway / Ma

看过 想看 没兴趣

推荐无处不在—淘宝

我的淘宝

商品页面

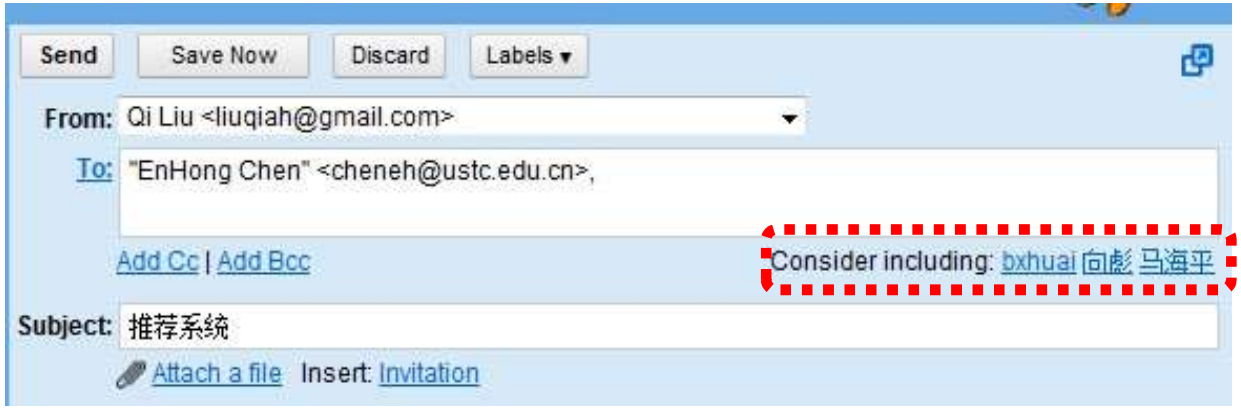
推荐无处不在—Amazon

背景

- 推荐无处不在


Google
recoomender
recommender
recommender systems
recommender salutation
recommender systems in e-commerce
recommender systems 2009
recommender systems research
recommender algorithm
recommender systems 2010
recommender systems conference
recommended synonym
Google Search
I'm Feeling Lucky


为您推荐相关宝贝:
侦探小说
随遇而安
中国史
挺这个靠谱历史
中国通史


Send
Save Now
Discard
Labels
From: Qi Liu <liuqiah@gmail.com>
To: "EnHong Chen" <cheneh@ustc.edu.cn>,
Add Cc | Add Bcc
Consider including: bxhuai 向彪 马海平
Subject: 推荐系统
Attach a file
Insert: Invitation

面向推荐算法的竞赛层出不穷

首届全国大学生数据挖掘邀请赛

面向推荐算法的竞赛层出不穷

面向推荐算法的竞赛层出不穷

阿里移动推荐算法	已结束	Alibaba Group 阿里巴巴集团	2015/07/01	¥300000	7186

MobileDM&HuMoComp 2015: The First International Workshop on Mobile Data Mining & Human Mobility Computing(ICDM 2015)

TIANCHI天池 首页 天池大赛 数据实验室 天池科学家 互动 御膳房

Organizers

Workshop Chairs:

- Rong Jin (iDST, Alibaba Group)
- Yixin Chen (Washington University)
- Qi Liu (University of Science and Technology of China)
- Zhongyi Liu (Alibaba Group)
- Nicholas Jing Yuan(Microsoft Research Asia)
- Rui Zhang(University Of Melbourne)
- Kai Zheng(University Of Queensland)

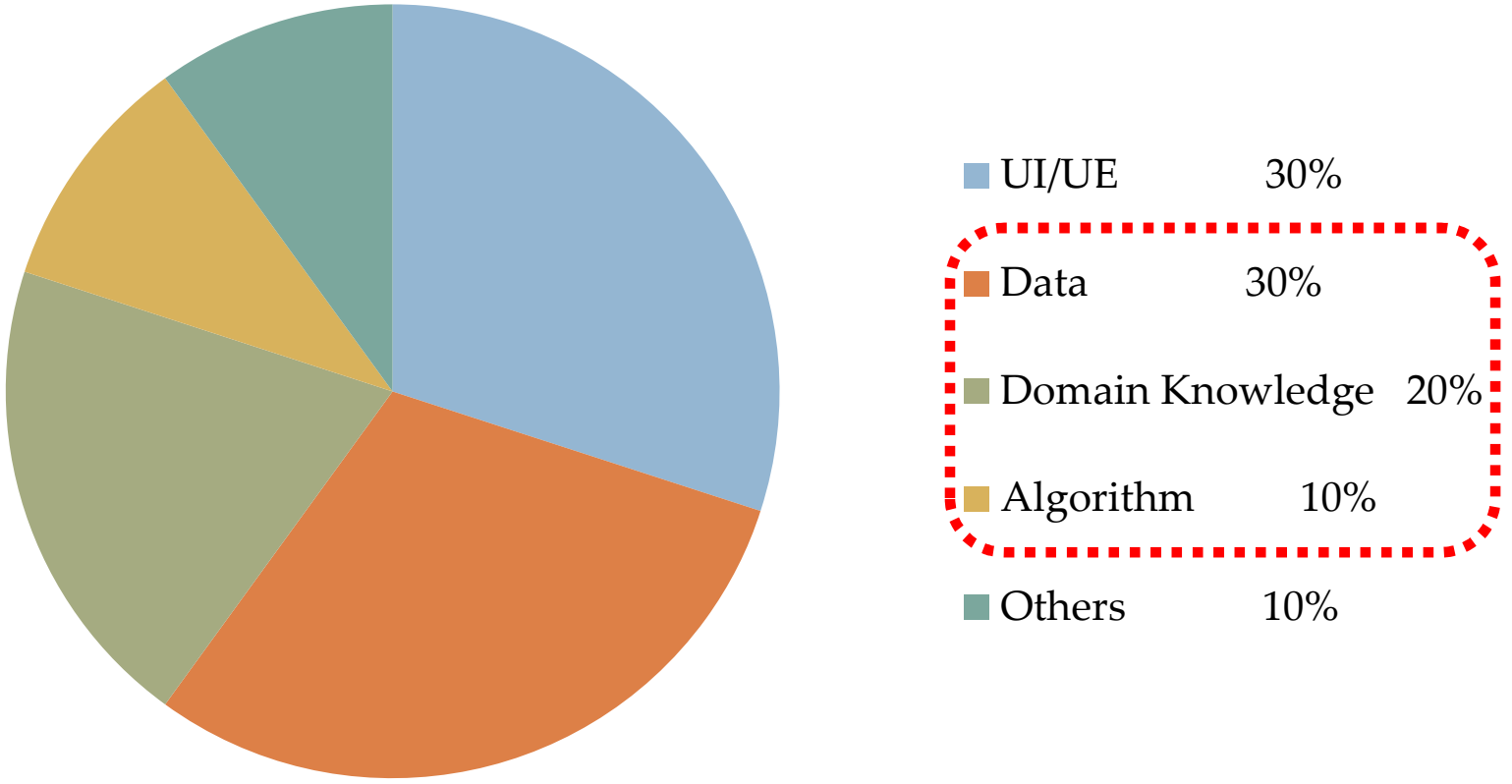
面向推荐算法的竞赛层出不穷

主要内容

- 什么是推荐系统
  - 背景、定义、应用场景
- 推荐方法概述
  - 兴趣建模
  - 推荐算法设计
  - 推荐结果的评估
- 案例学习
  - 基于用户兴趣扩展的个性化推荐方法
  - 面向推荐系统的纠结心理挖掘
- 小结及未来的路
- 资料推荐

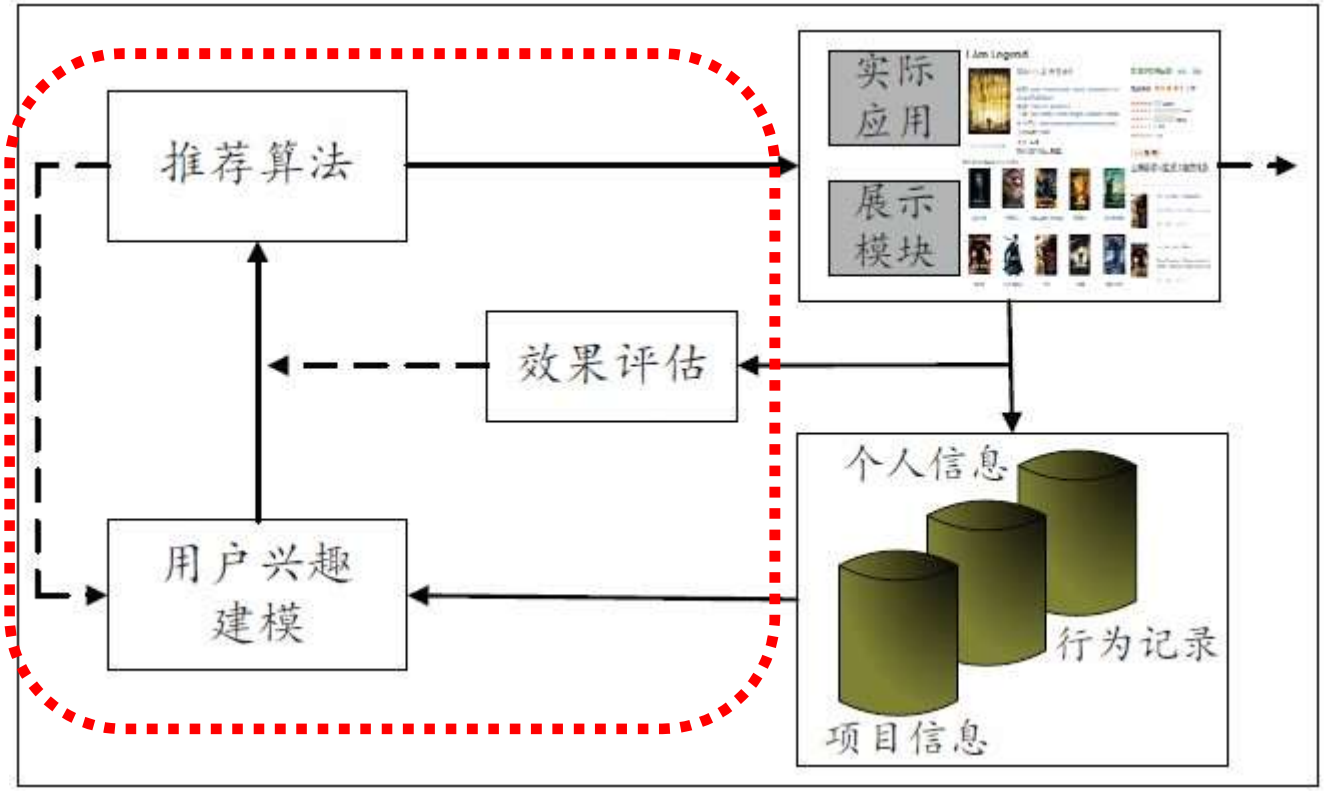
背景及相关工作

- 推荐系统
  - 首先是一个“System”



背景及相关工作

- 推荐系统框架



相关工作

- 推荐方法分类


推荐算法
设计
混合算法
协同过滤
基于内容算法
★★★★★力荐
评分预测
排序预测
结果评估
非个性化兴趣
模型
个性化兴趣
模型
融合情境信息的
个性化
兴趣
建模

主要内容

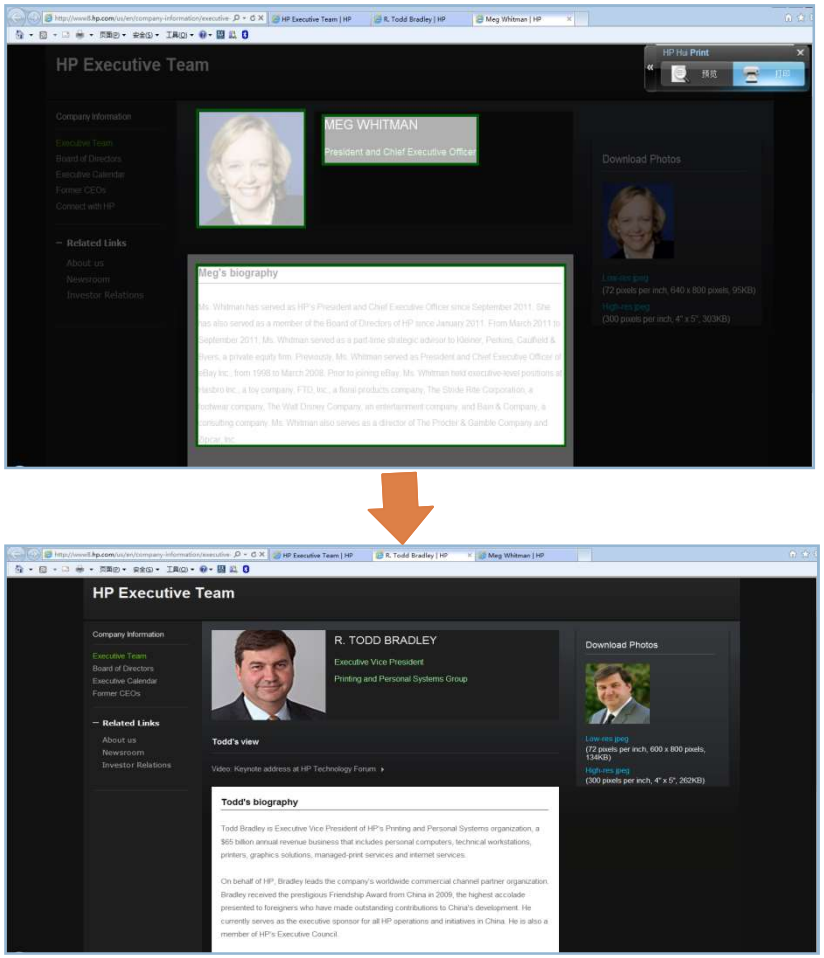
- 什么是推荐系统
  - 背景、定义、应用场景
- 推荐方法概述
  - 兴趣建模
  - 推荐算法设计
  - 推荐结果的评估
- 案例学习
  - 基于用户兴趣扩展的个性化推荐方法
  - 面向推荐系统的纠结心理挖掘
- 小结及未来的路
- 资料推荐

兴趣建模

- 非个性化兴趣模型
  - 无法识别单个用户的ID
    - 群体智慧

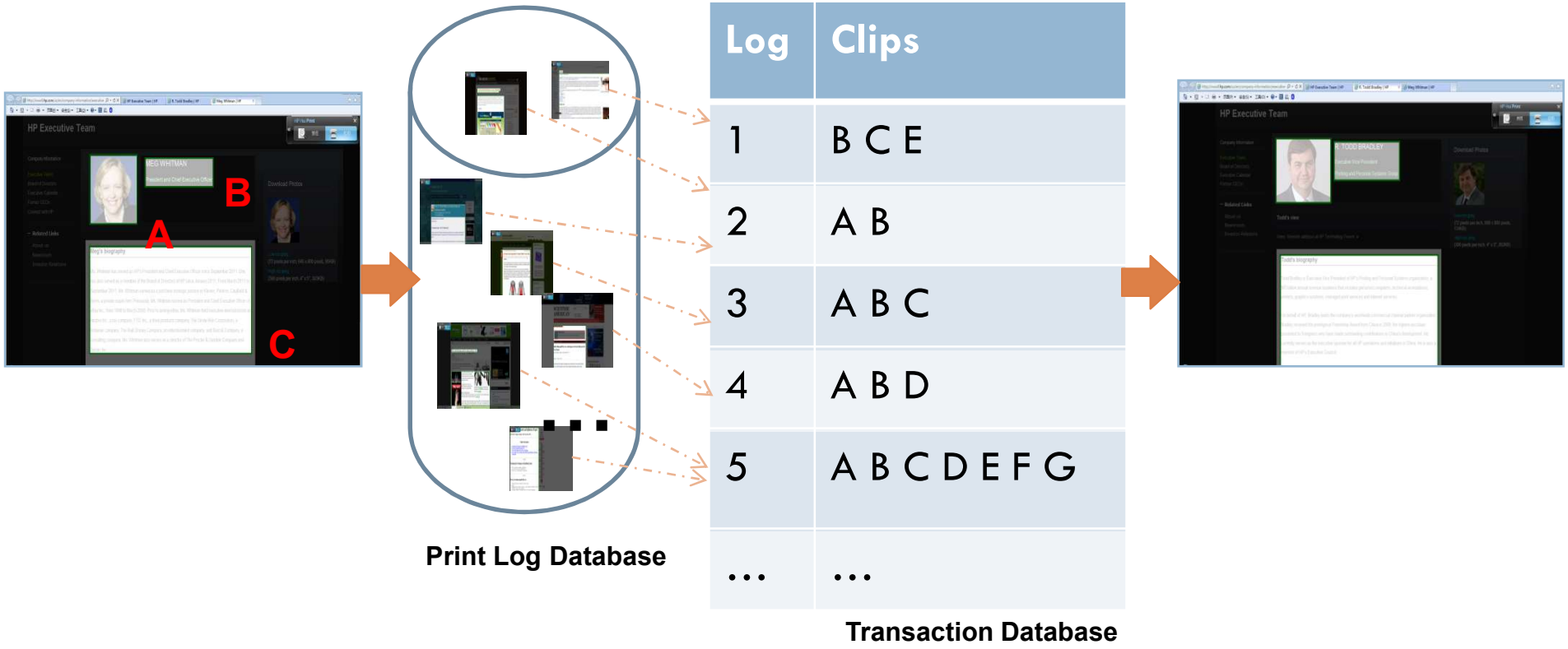
Google

recoomender
recommender
recommender systems
recommender salutation
recommender systems in e-commerce
recommender systems 2009
recommender systems research
recommender algorithm
recommender systems 2010
recommender systems conference
recommended synonym
Google Search I'm Feeling Lucky



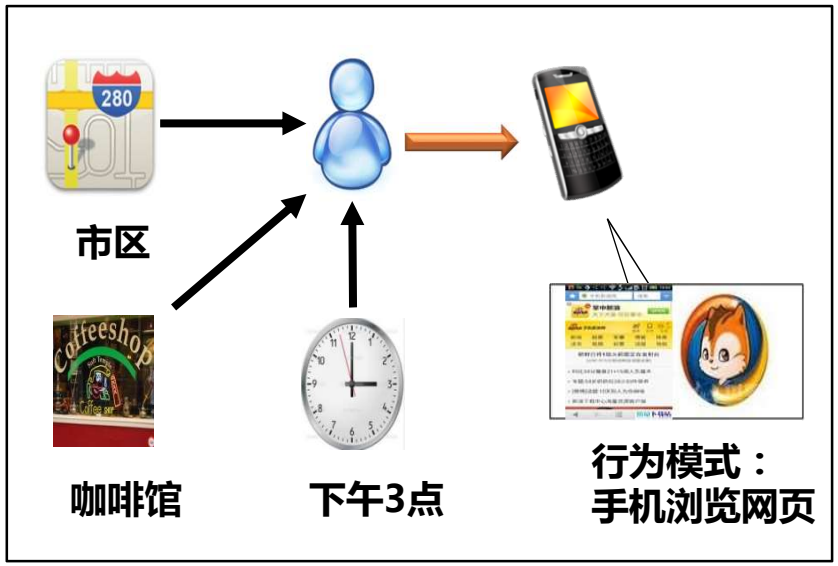
兴趣建模

- 非个性化兴趣模型
  - 无法识别单个用户的ID



兴趣建模

- 非个性化兴趣建模
- 关联分析 ----发现经常出现的事物、行为、现象
  - 购买A的顾客还会购买。。。
  - 搜索“讯飞”的用户还会经常搜索“科大讯飞”
  - 移动用户在下午五点经常听歌还是玩游戏？
  - 下载某类音乐的用户通常有哪些特点？



用户行为与所处环境的关联

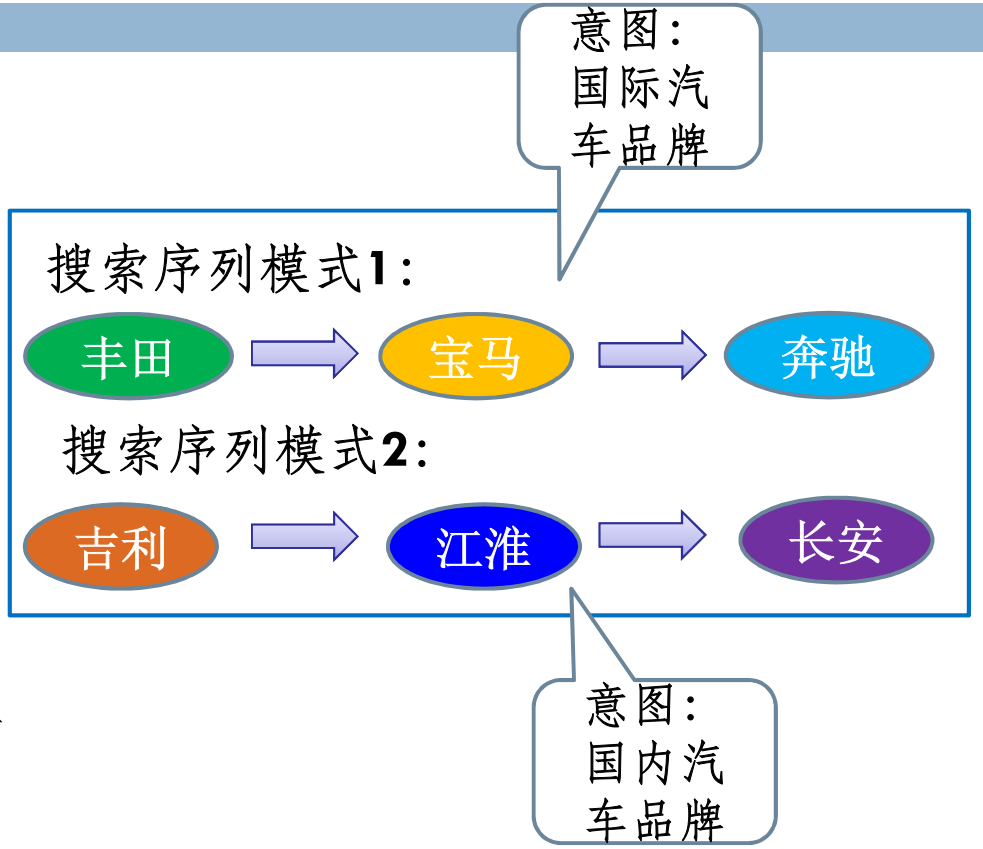
{(Is a holiday?:Yes),
(time range: PM3:00-4:00),),
(location: Urban)}

⇒ *Surfing on the Internet*

兴趣建模

- 频繁序列模式挖掘

SID	Search Session
1	丰田➔雷诺➔宝马➔奔驰
2	宝马➔奔驰➔法拉利
3	本田➔丰田➔通用➔宝马
4	吉利➔奇瑞➔长城➔江淮
5	比亚迪➔吉利➔江淮➔长安
6	长城➔江淮➔华泰➔长安



- 类似于关联模式挖掘的算法
  - PrefixSpan
- 隐马尔科夫模型、CRF
  - 将搜索意图理解为隐状态
  - 建模搜索历史与当前意图之间的序列模式

兴趣建模

- 个性化兴趣模型
  - 利用项目列表来表示用户兴趣
    - 加权 or 不加权
    - 数据异常稀疏

	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6
User-1	4	*	2	5	*	*
User-2	3	2	1	*	*	3
User-3	*	2	*	3	*	4
User-4	*	3	3	5	4	*
User-5	5	*	3	4	*	*

兴趣建模

- 个性化兴趣模型

	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6
User-1	4	*	2	5	*	*
User-2	3	2	1	*	*	3
User-3	*	2	*	3	*	4
User-4	*	3	3	5	4	*
User-5	5	*	3	4	*	*

Three kinds of data: users' data, items' data, and ratings' data.

- Name
- Job
- Like
- Address
- ……

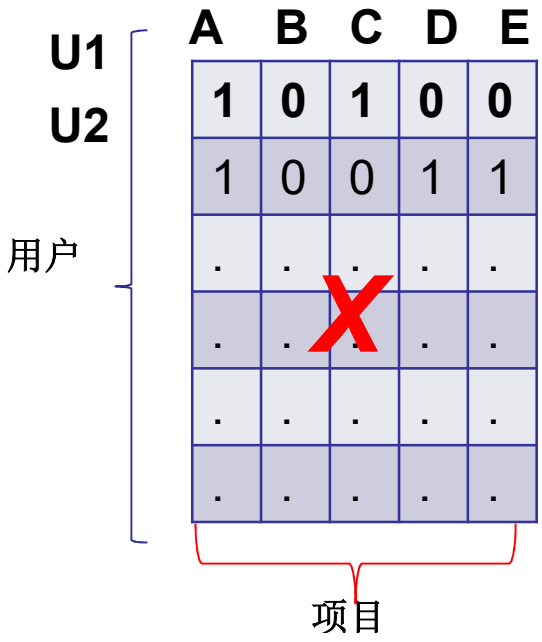
- Name
- Tag
- [content]
- [category]
- [time]
- ……

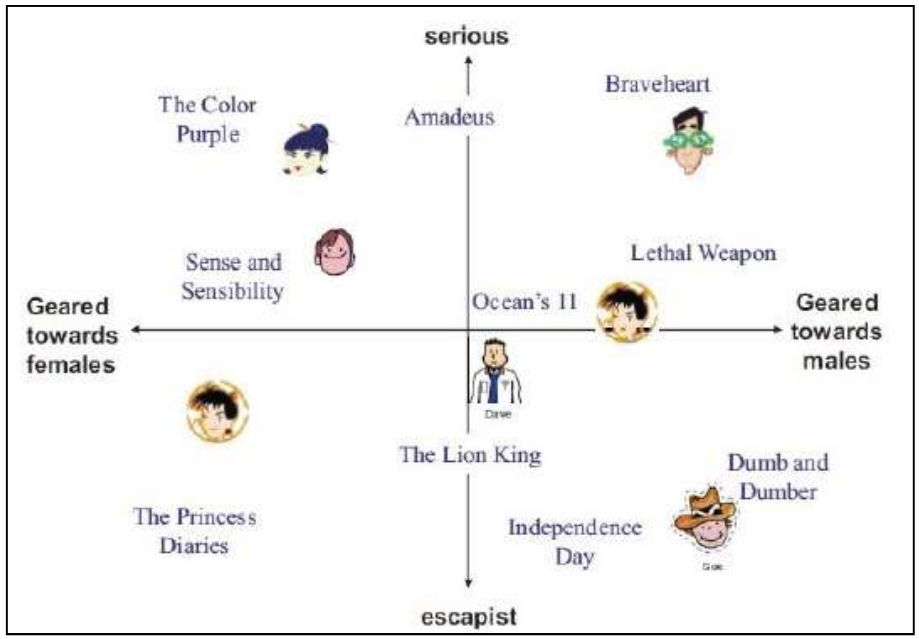
- Value
- User-id
- Item-id
- [time]
- ……

兴趣建模

- 个性化兴趣模型
  - 低秩分解
    - 评分预测
      - 1-5分
    - 学习排序
      - 1购买，0未购买

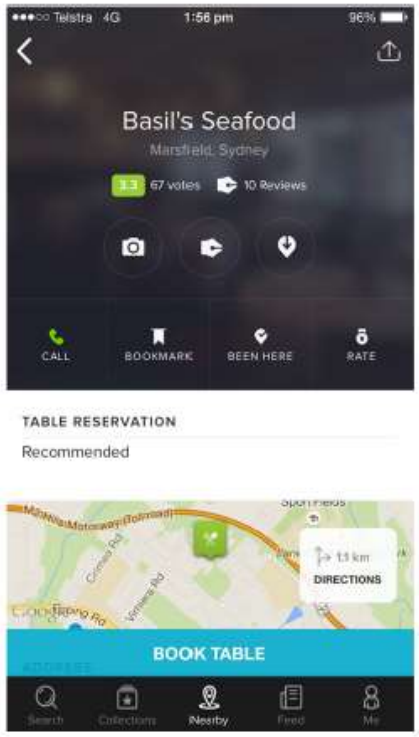
	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6
User-1	4	*	2	5	*	*
User-2	3	2	1	*	*	3
User-3	*	2	*	3	*	4
User-4	*	3	3	5	4	*
User-5	5	*	3	4	*	*

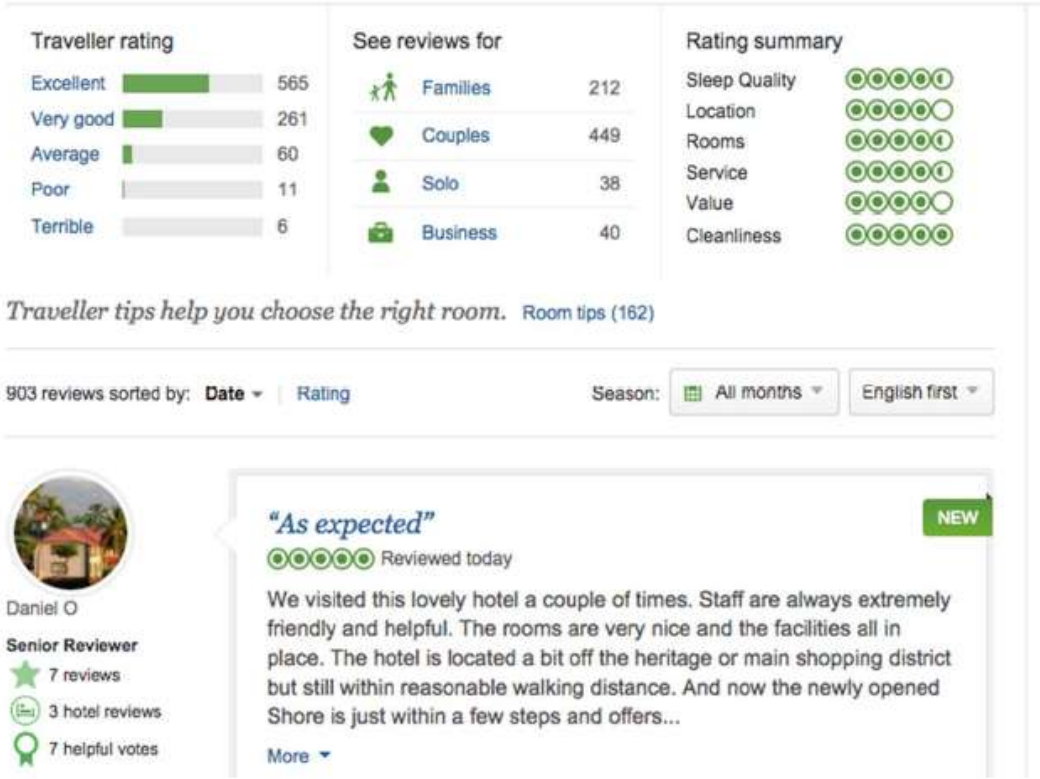




兴趣建模

- 显式信息(**Explicit Information**)
  - 用户手动添加，系统直接从用户输入获得
    - 评分、点赞、评论等





兴趣建模

- 隐式信息(Implicit Information)
  - 自动生成，要么是和系统的交互信息，要么是模型训练所得
    - 例如点击、浏览等日志行为
      - 还如鼠标轨迹、浏览器缓存等等

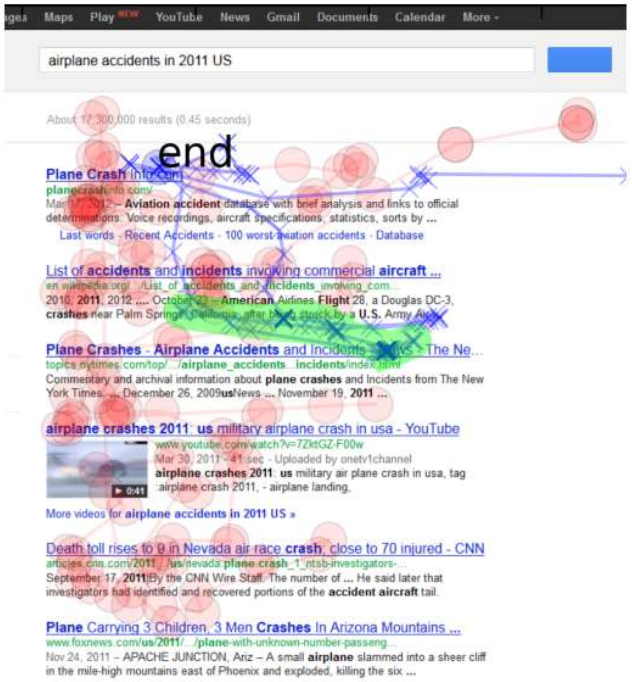


Figure 1: An example automatically discovered motif from mouse

Table 1: A toy example of the customer behaviorial records.

UserId	ItemId	CategoryId	Action	Timestamp
U_1	a	C_1	Click	2014-07-08 20:05:20
U_1	b	C_1	Click	2014-07-08 20:06:40
...	...	...	...	...
U_1	a	C_1	Cart	2014-07-08 20:13:55
U_1	b	C_1	Collect	2014-07-08 20:14:20
U_1	b	C_1	Buy	2014-07-08 20:14:38
U_2	f	C_2	Click	2014-07-09 10:21:13
U_2	f	C_2	Buy	2014-07-09 10:21:20

兴趣建模

- 融合情境信息的个性化兴趣模型
  - 基于情境信息的用户行为切分
    - 移动情境

●当前时间
●用户位置
●移动速度
●周围噪声

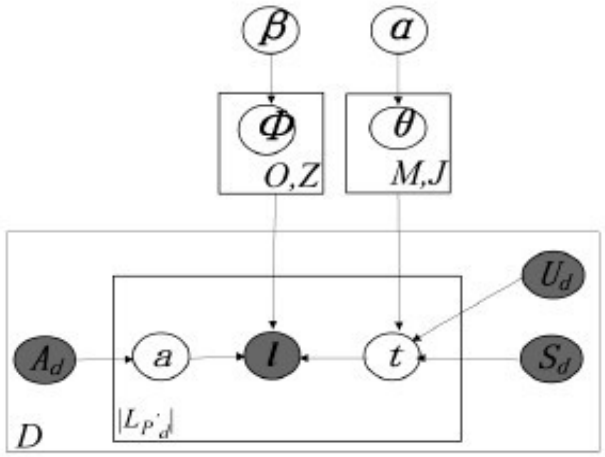
。。。

兴趣建模

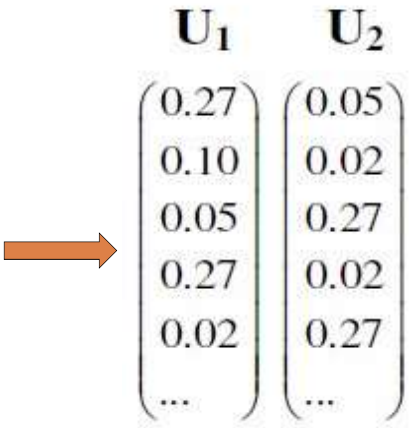
- 融合情境信息的个性化兴趣模型
  - 基于情境信息的用户行为切分
    - 旅游情境

兴趣建模

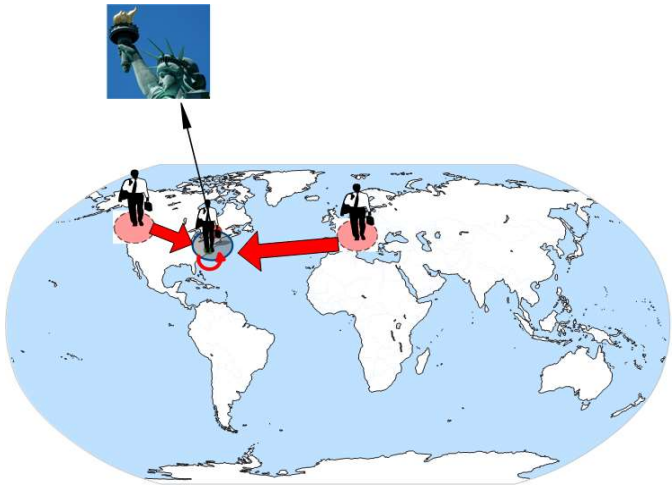
- 融合情境信息的个性化兴趣模型
  - 基于情境信息的用户行为切分
    - 旅游情境



情境感知的用户兴趣挖掘模型

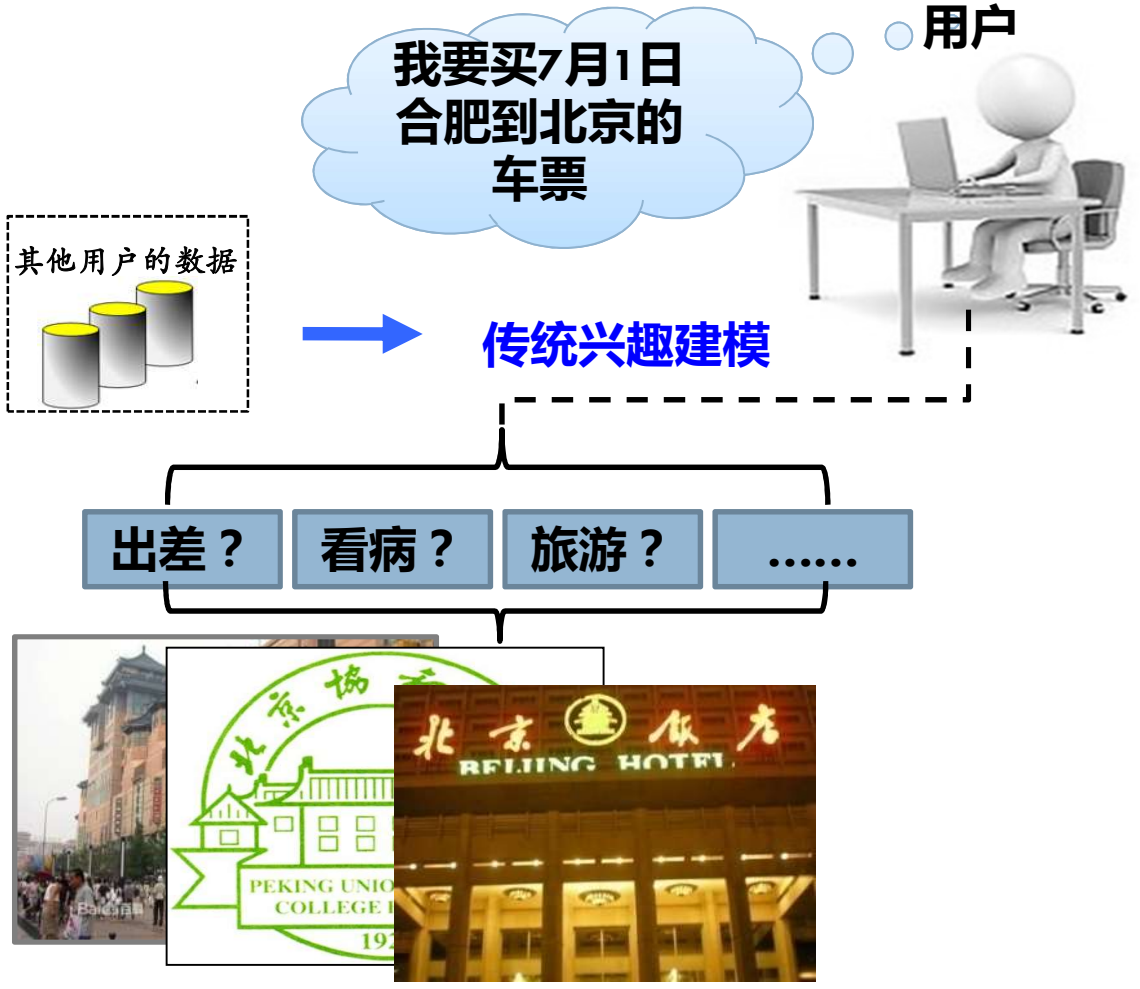


用户兴趣表示



兴趣建模

- 融合情境信息的个性化兴趣模型



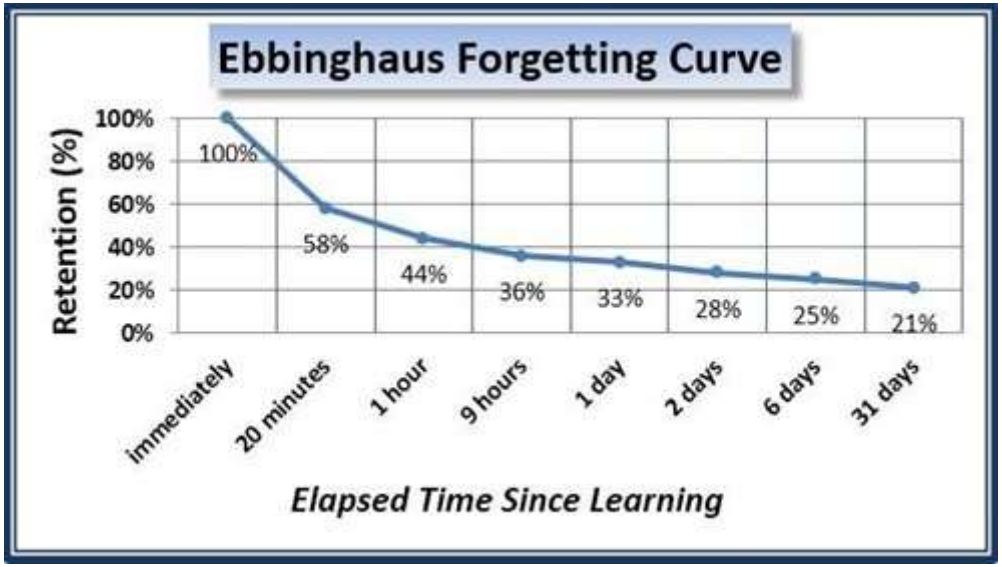
- 无情境识别
- 不区分用户在不同情境中的需求

兴趣建模

- 融合情境信息的个性化兴趣模型

兴趣建模

- 融合能力的个性化兴趣模型
  - 基于能力的用户画像
    - 学习场景
      - 知识点：数学、物理、词汇
      - 技能：记忆力、逻辑、演算

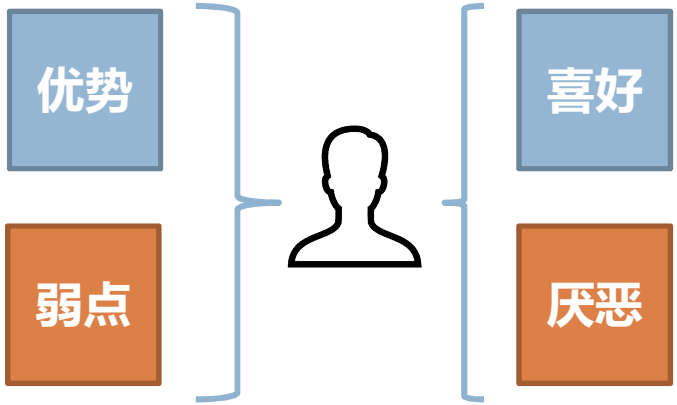


兴趣建模

- 融合能力的个性化兴趣模型
  - 基于能力的用户画像
    - 竞技场景
      - 竞技技能：速度、力量、防守
      - 竞技意识：协防、突袭、反击

兴趣建模

- 融合能力的个性化兴趣模型
  - 基于能力的用户画像
    - 能力 vs 兴趣

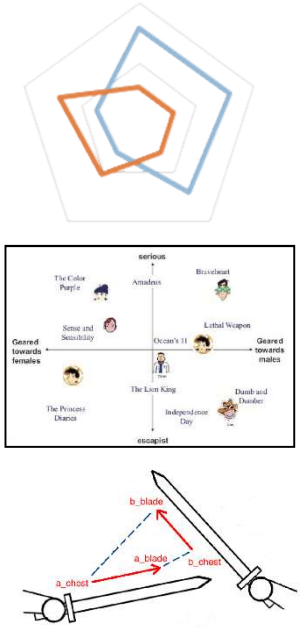


    - 匹配推荐

用户能力建模

用户兴趣建模

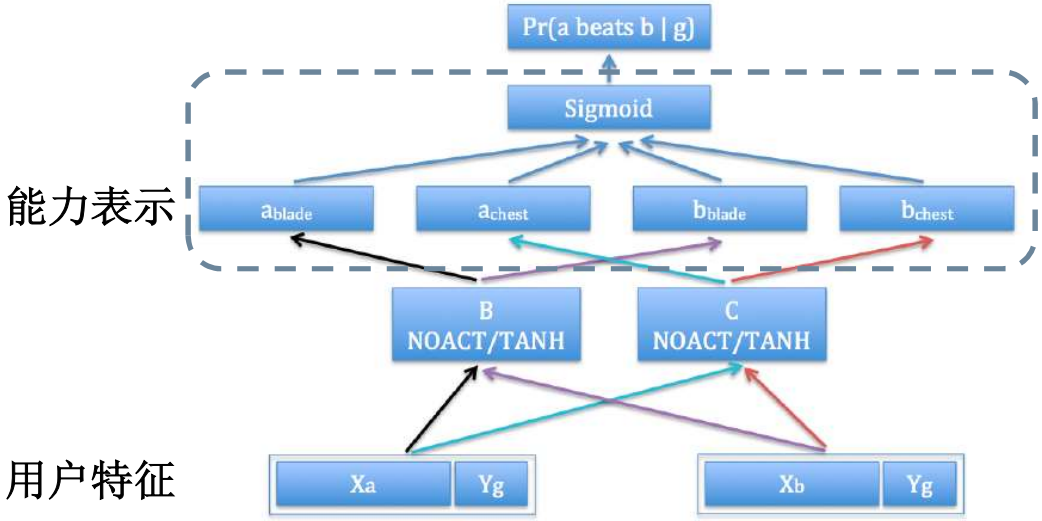
匹配推荐



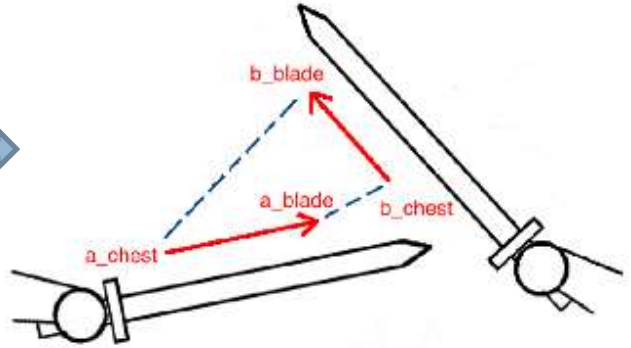
匹配合适的项目或者合适的对手

兴趣建模

- 融合能力的个性化兴趣模型



基于对抗的用户能力表示模型



Blade：优势or锋芒
Chest：弱点or软肋

“锋芒-软肋”对抗模型

$$\Pr(a \text{ beats } b) \propto ||b_{blade} - a_{chest}||_2^2 - ||a_{blade} - b_{chest}||_2^2$$