

领域知识驱动的个性化推荐方法

陈恩红 教授

cheneh@ustc.edu.cn

2016年12月6日

1

大数据时代背景及现状

2

基于认知的教学个性化推荐

3

情景感知的移动用户推荐

4

金融领域带风险约束的推荐

5

结合社交因素的用户推荐

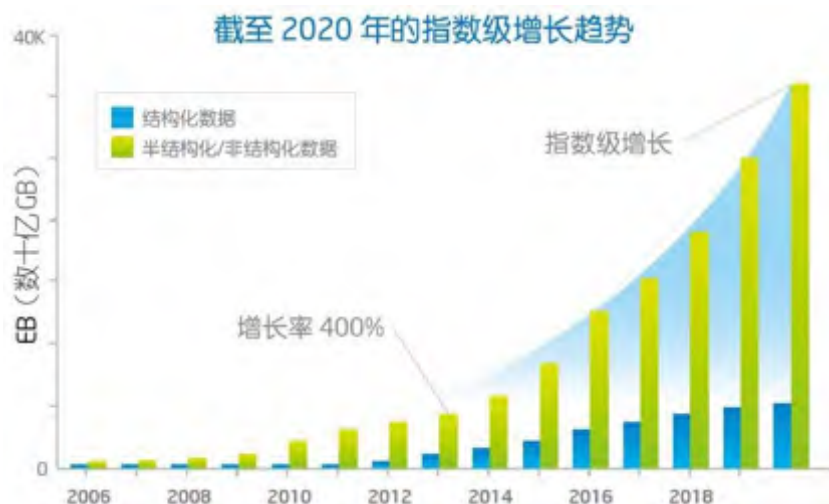
6

总结与展望

大数据与大数据时代

■ 大数据时代：数据爆炸增长的时代

- 数据量大、类型多、价值密度低、速度快时效高、数据在线
- 各行业急速增长的数据量为推荐系统的应用奠定了基础



■ 大数据从“概念”走向“应用”

- 分析大数据成为科研的核心任务
- 解读大数据成为决策的基本依据
- 应用大数据成为民生的重要助力



推荐系统蕴含巨大价值

■ 推荐系统成为连接数据特征与用户需求的桥梁

- 普通用户难以直接从大数据中获取所需信息
- 推荐系统将大数据从单纯的数据层面转化到用户可以理解的信息层面，满足客户的需求



■ 推荐系统为企业带来**巨大价值**

- 2016年“双十一”淘宝交易额破千万
- 今日头条app人均日使用时间超40分钟
- UC浏览器个性化推荐月活跃用户超过3.3亿



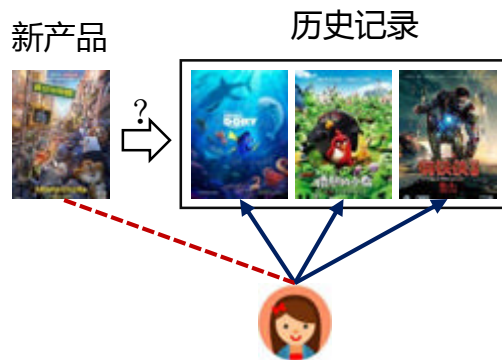
我们正在离开信息的时代，进入推荐的时代。

——Chris Anderson in *The Long Tail*

传统推荐系统常用方法

■ 基于内容的推荐算法

根据用户**偏好记录**来给用户推荐内容相似的新项目

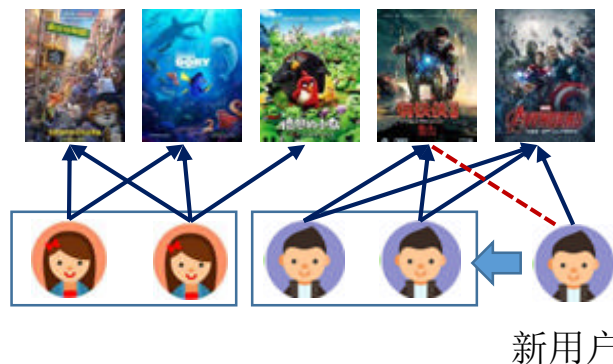


优点：不存在稀疏性和冷启动问题

缺点：可能会重复推荐

■ 协同过滤推荐算法

根据用户**最近邻的偏好**记录给用户作出推荐



优点：可处理复杂的非结构化对象

缺点：稀疏性和冷启动问题

■ 混合推荐算法

结合基于内容的特性和协同过滤的特性进行推荐

传统推荐系统面对的困难

泛化的数据分析结果**难以精准描述**用户在**特定应用场景**中的信息需求。

- 大数据的应用模式不断拓展
 - 专业性不断增强
 - 很多应用新问题**无法使用泛化模型**进行统一建模
- 数据的多源异构性明显
 - 数据量**大而复杂、来源分散**
 - 需要依托专业知识进行有效整合



推荐系统将扩展至更多领域

■ 大数据时代推荐系统向更多领域扩展

- 传统行业开始将业务转移至“线上”
- 移动软件已经融入人们的日常生活
- 各个领域都积累了大量的数据资源



电子
商务

医疗
卫生



BIG
DATA

推荐

教育
评估



移动
用户



社交
应用



智慧
城市



领域知识驱动推荐技术

- 泛化的推荐结果无法满足具体领域中特定的用户需求



- 不同领域独有的知识体系和知识结构
- 不同行业独有的评价准则

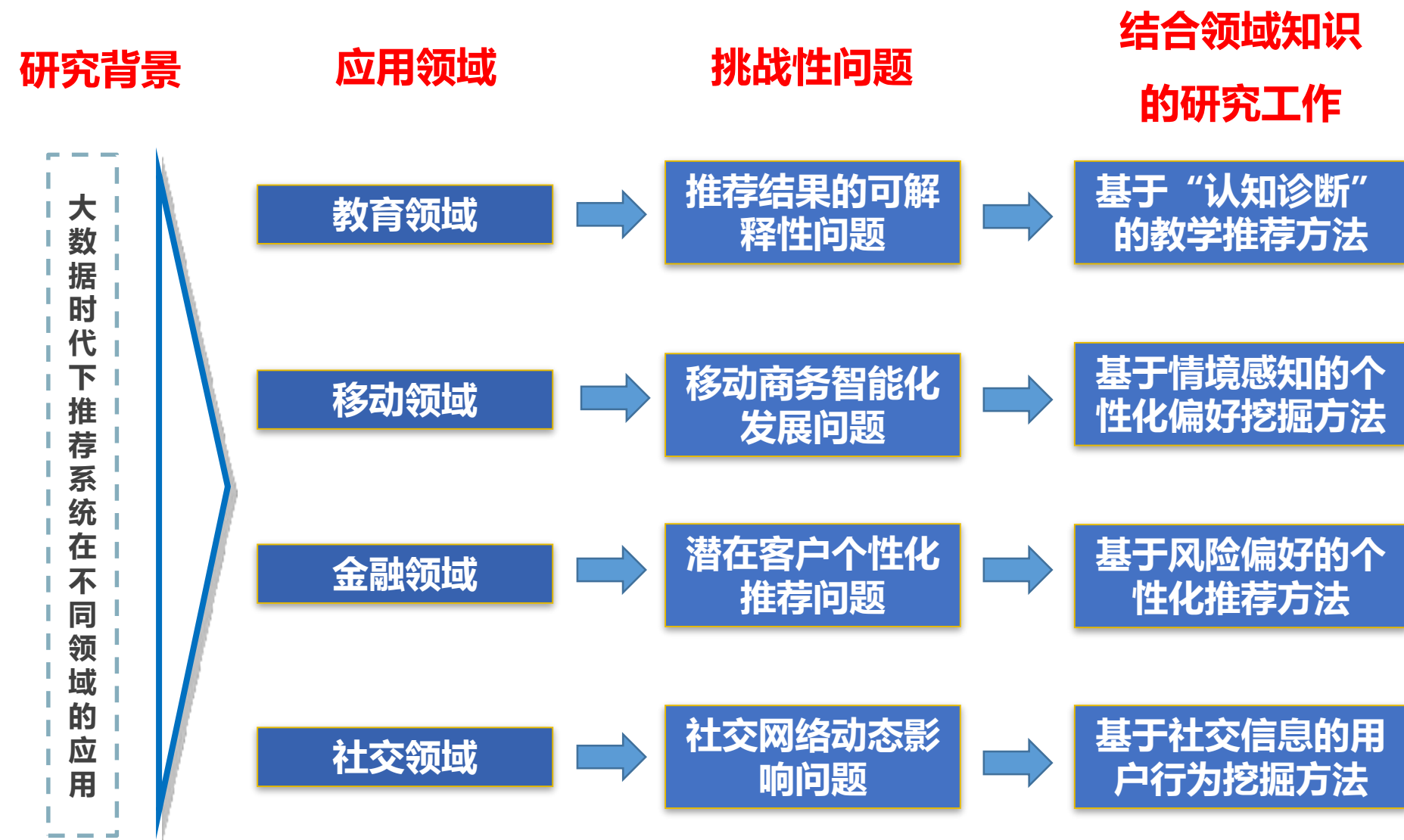
新的应用领域

- 教育领域
- 移动领域
- 金融领域
- 社交领域
- ...

- 结合领域知识定制推荐系统



推荐系统在各领域的应用案例



1

大数据时代背景及现状

2

基于认知的教学个性化推荐

3

情景感知的移动用户推荐

4

金融领域带风险约束的推荐

5

结合社交因素的用户推荐

6

总结与展望

教育领域的推荐目标

教育领域所面临的问题：



- 学生对自己的学习状态没有直观的认识

认知诊断



- 学生作业多、课业量繁重

试题推荐



- 班级教学中，老师对于班级内的学生难以顾全

学习分组推荐

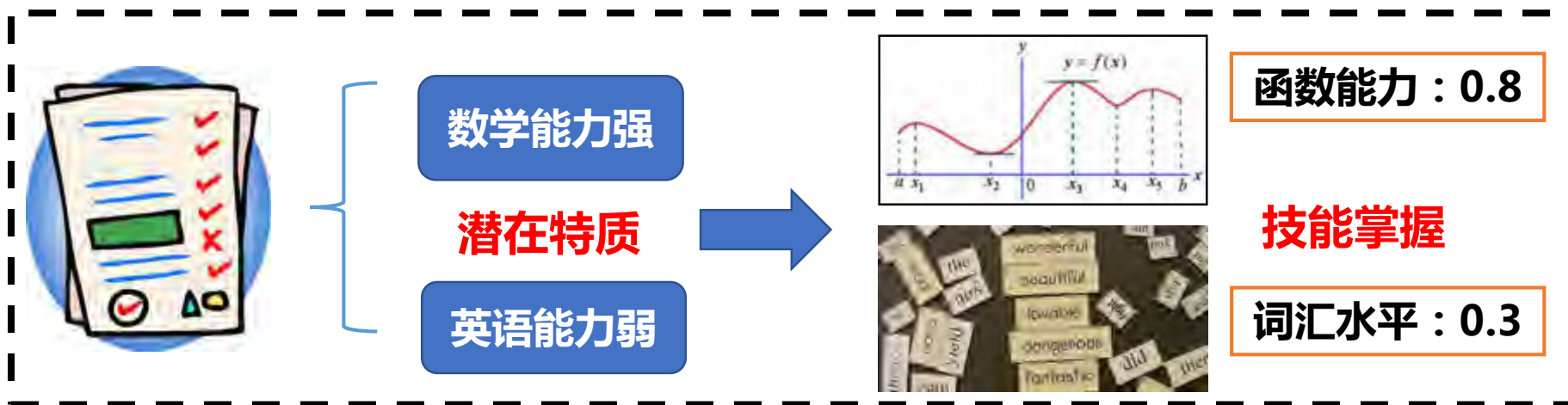
教育推荐的特点和所面临的挑战

教育推荐的特点：

- 有特殊的领域知识（试题知识点、学生认知情况等），教育推荐需要同“**认知诊断**”等教育领域方法相结合
- 用户（学生、教师等）对于推荐结果的可解释性要求较高

所面临的挑战：

- 数据挖掘技术**如何同教育学的领域知识相集合**？
- 如何在不同的教学推荐中引入领域知识，使得推荐结果更精确、可解释性更高，从而提高用户（学生、教师等）的接受程度？



如何评价一个学生的能力？

传统的评价方法：

- 基于学生成绩的评价方法
- 基于学生排名的评价方法



问题

- 评价粒度粗
- 结果单一
- 没有个性化分析



认知诊断评价方法：

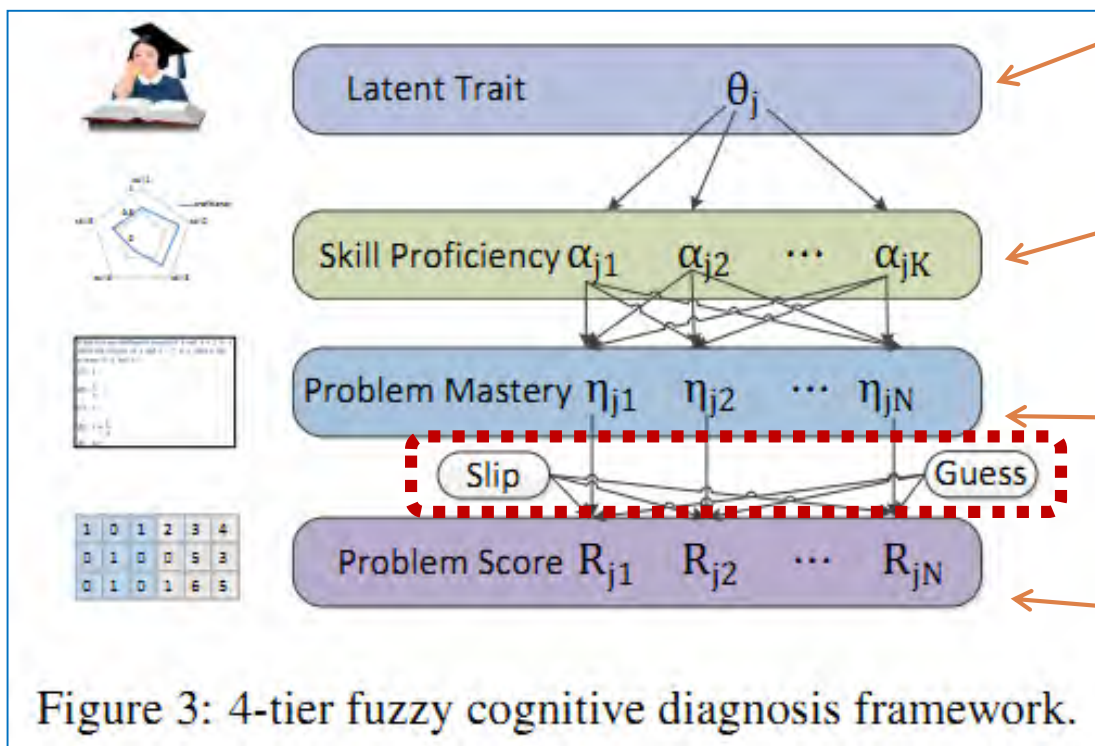
- 结合答题数据和教育心理学方法
- 能够分析学生的知识状态



“认知诊断”的学生建模方法

框架：模糊认知诊断模型(四层生成模型)

- 概率化学生技能掌握程度
- 结合模糊理论，分别建模客观题与主观题的技能反映模式
- 考虑学生不确定因素（粗心”和“猜测”）



学生潜在特质/能力

学生知识点掌握程度

学生题目掌握程度

学生题目具体得分

学生画像建模层级

诊断结果用于教育领域个性化推荐

诊断结果的应用

- 教育领域的应用场景**复杂多样**。
- 如何在不同应用场景的目标之间，**引入学生认知诊断的结果**，和大数据有效的结合，进行更合理、更高效的个性化推荐？

场景1：学生个体学习



学生个性化试题推荐

场景2：班级群体学习



班级学生分组推荐

场景1：学生个体学习

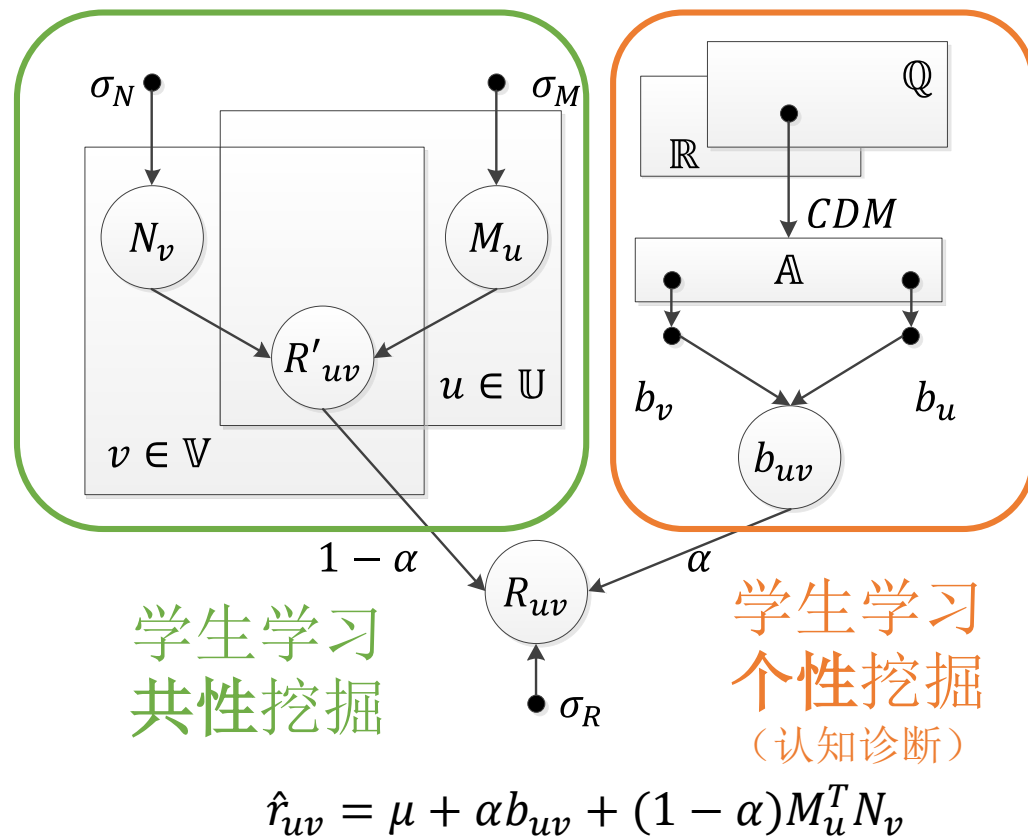
如何从海量的试题中自动筛选出适合每个学生的题目？

- 已知学生的个性化知识点掌握程度，和现有的得分预测方法结合，使推荐结果更加准确。

基于“认知诊断”的得分预测

- 用几何平均计算学生在试题所需技能上的平均技能掌握程度，得到矩阵A。
- 从矩阵A中提取特征 b_{uv} 作为PMF的先验信息。将得到新的推荐模型PMF-CD

$$b_{uv} = b_u + b_v, \text{ 其中: } \begin{cases} b_u = \frac{1}{V} * \sum_{i=1}^V A_{ui} \\ b_v = \frac{1}{U} * \sum_{i=1}^U A_{iv} \end{cases}$$

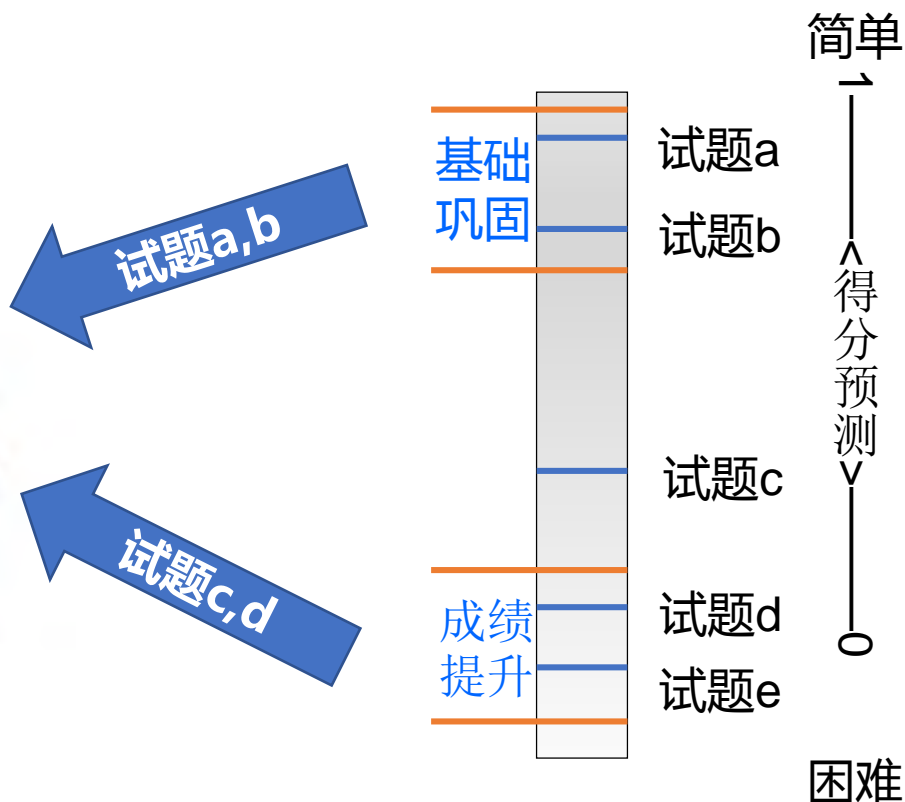


基于“认知诊断”的试题推荐

针对不同场景的学生个性化试题推荐

- 在试题推荐中，不是推荐越难的试题越好、也不是推荐越简单的试题越好，这和传统的推荐系统有较大的差异。
- 根据试题相对于学生的难度，从试题库中为学生推荐个性化试题。

根据不同的学习目标，向学生提供不同难度的试题



场景2：班级群体学习

班级群体学习推荐

- 教育大数据对学生的分析除了可以应用于针对每个学生的个性化试题推荐之外，还可以为线下的教学活动进行推荐。
- 根据认知诊断所得到的**学生个性化知识点掌握程度**，为同班级的学生进行协同学习分组推荐。

协同学习分组

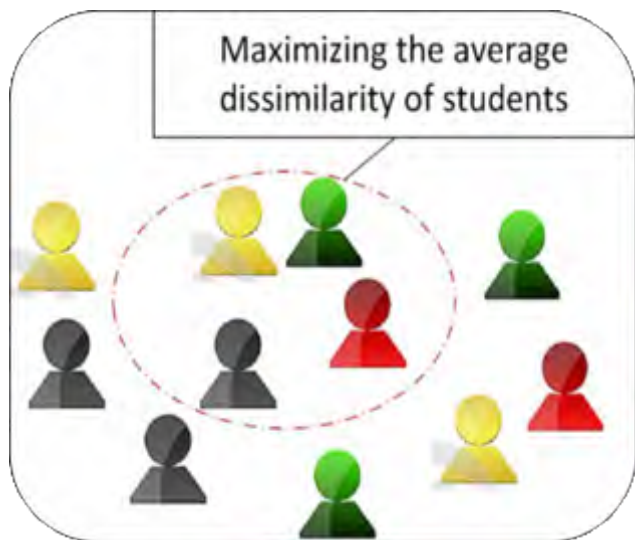
- 通过**学习小组**的形式组织学生进行学习的一种教学应用
- 目标：将学生进行分组，将班内所有学生分入若干个组中，组内学生通过相互交流、讨论，达到**学业水平提升**



基于“认知诊断”的协同学习分组

(A) 基于学生差异的分组算法UKB:

- 分组目标：将最大化每个组内学生的平均异质性
- “认知诊断”结果相似的学生分在不同的组，使得每个分组中的学生可以优势互补。



(A) 基于异质性的目标

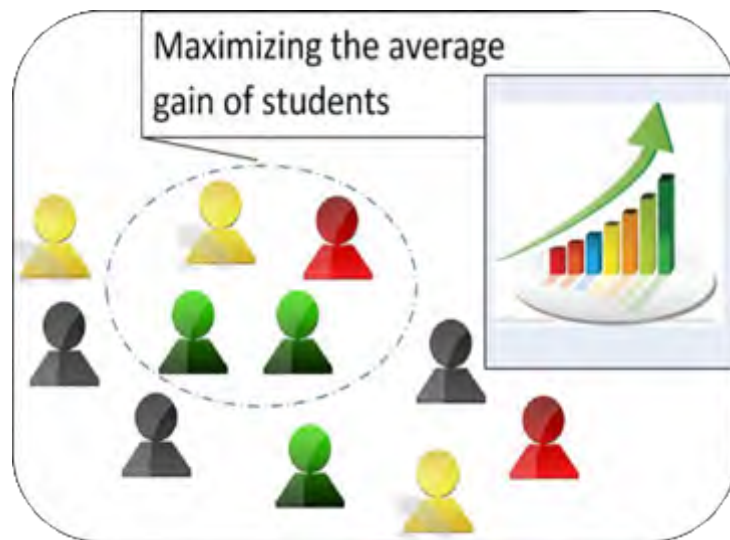
(B) 基于收益的学生分组算法BGB:

- 分组目标：最大化所有学生的平均收益
- 学生收益：受学生技能掌握和组内leader技能掌握共同影响

$$\text{学生收益} = (a_{kj} - \alpha_{ij}) \cdot \alpha_{ij}$$

组内最好学生技能掌握程度

学生本人技能掌握程度



(B) 基于收益的目标

根据学生的考试情况生成立体成绩报告，帮助学生了解自己的学习状态，并向学生推荐个性化练习题。

1



我考得怎么样?
单科成绩概述、各班成绩对比

2



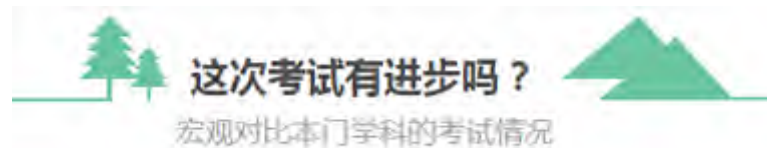
解析丢分题 **我的分数丢哪了?**
发现问题，分析问题，总结经验

3



知识点诊断 **考试知识点，我能知多少?**
了解考试知识点，把握自身掌握度

4



这次考试有进步吗?
宏观对比本门学科的考试情况

5



我该怎么办?
找到关键问题，听听智学小建议

在线试题推荐流程



同步学习任务

最新任务

- 5_3测试
- 测试包
- 练习包1
- 练习包2
- 练习包3
- 复测包

历史任务

现在立马学

简单, 很快能学会

关卡1: 数学归纳法

关卡2: 集合的含义

答题

基于认知诊断结果的个性化试题推荐

1

大数据时代背景及现状

2

基于认知的教学个性化推荐

3

情景感知的移动用户推荐

4

金融领域带风险约束的推荐

5

结合社交因素的用户推荐

6

总结与展望

移动商务大数据

- 2015年，全球智能手机用户比例超过全球人口的十分之一

- 传统的移动数据来源分散、关联性差

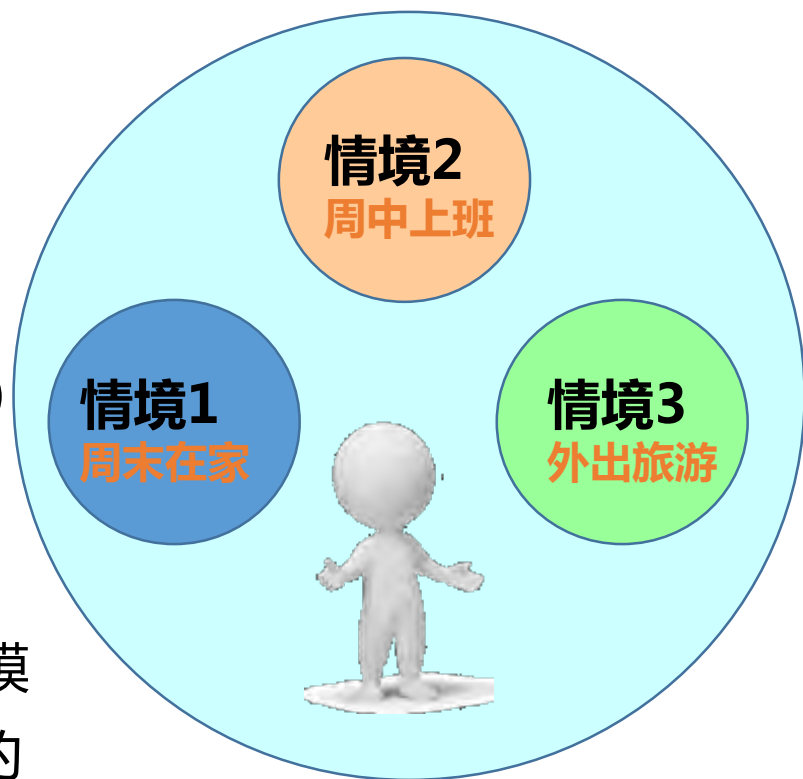
- 用户移动行为数据
- 地理信息数据

- 移动商务研究的新机遇：情境数据

- 用户属性（年龄、性别、身高等）
- 历史数据（消费记录、搜索上下文等）
- 环境数据（自然环境、社会关系等）

- 情境数据挖掘

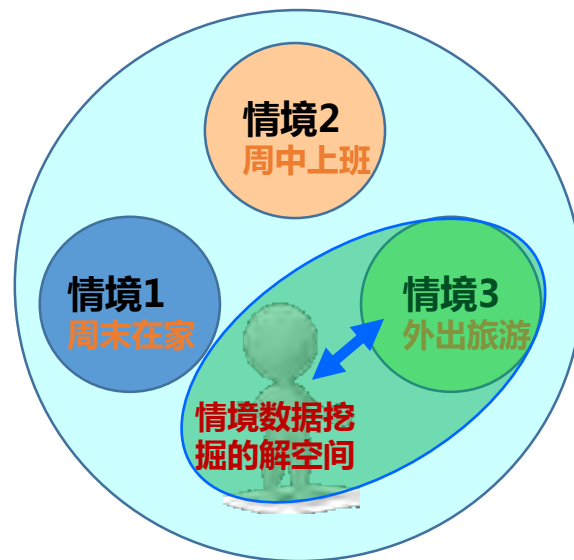
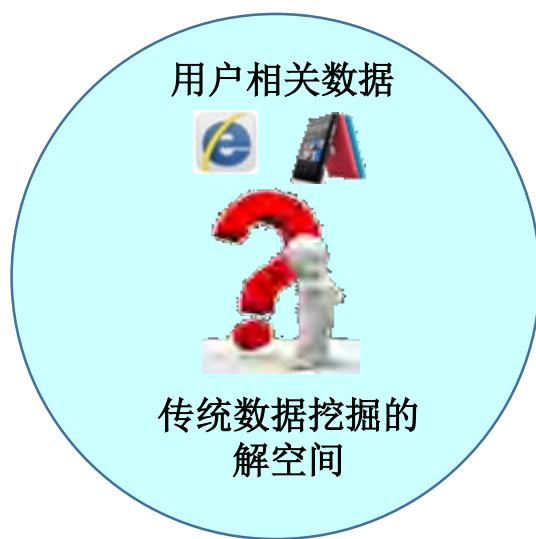
- 针对用户相关的情境数据进行情境建模和特征提取，挖掘出用户在不同情境下的特性或者行为模式的过程



基于情境感知的移动推荐

■ 情境数据挖掘的特点

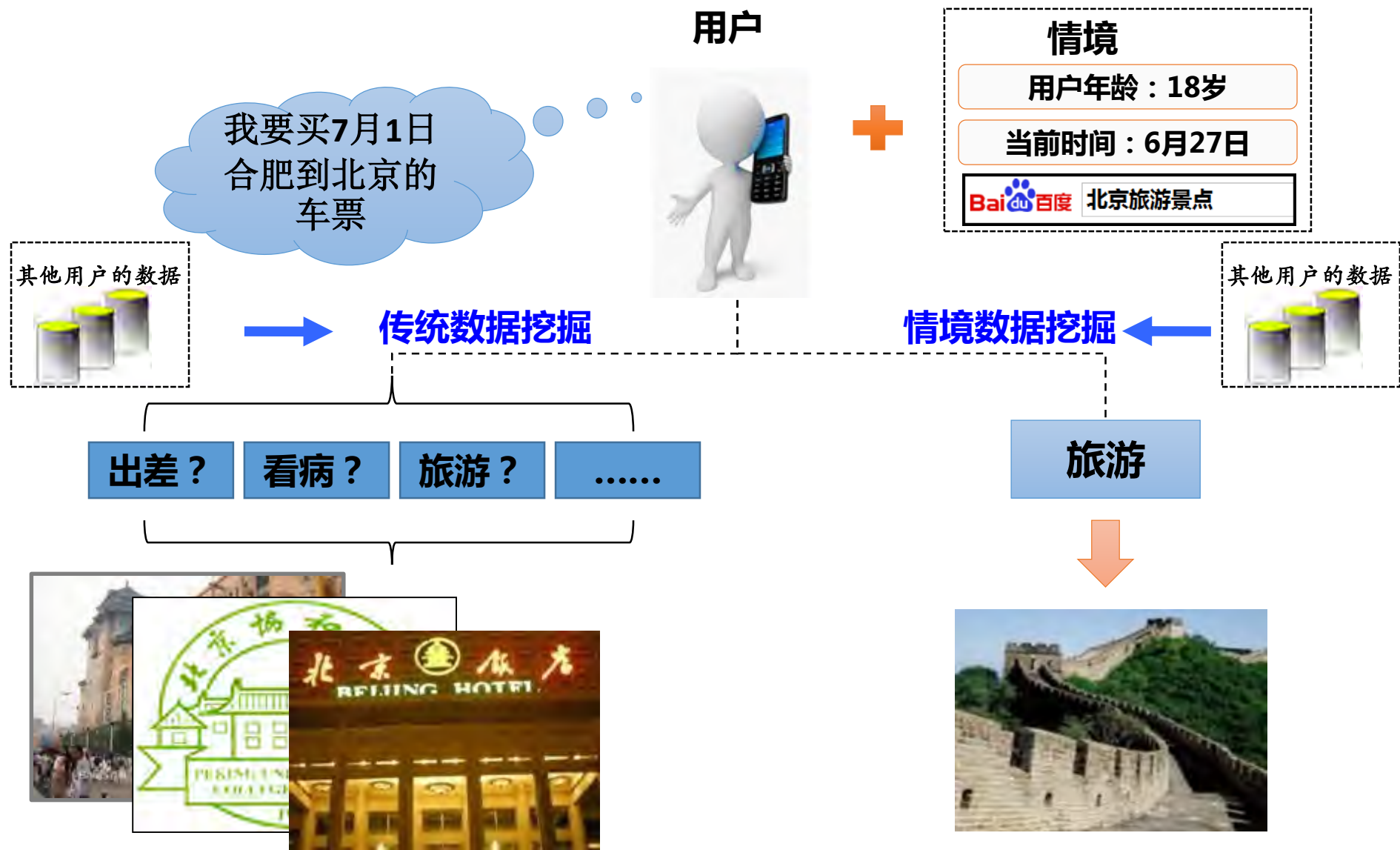
- 传统数据挖掘方法挖掘的解空间较大，结果难以满足用户需求
- 情境数据挖掘缩小解空间，对用户进行个性化推荐



■ 情境数据挖掘的挑战

- 用户情境的**动态化**
- 情境信息来源的**多样化**
- 情境特征的**复杂性**及不同特征间的**关联性**

基于情境感知的移动推荐



相似行为习惯的移动用户挖掘

相似用户挖掘的意义



移动用户情境数据的特点

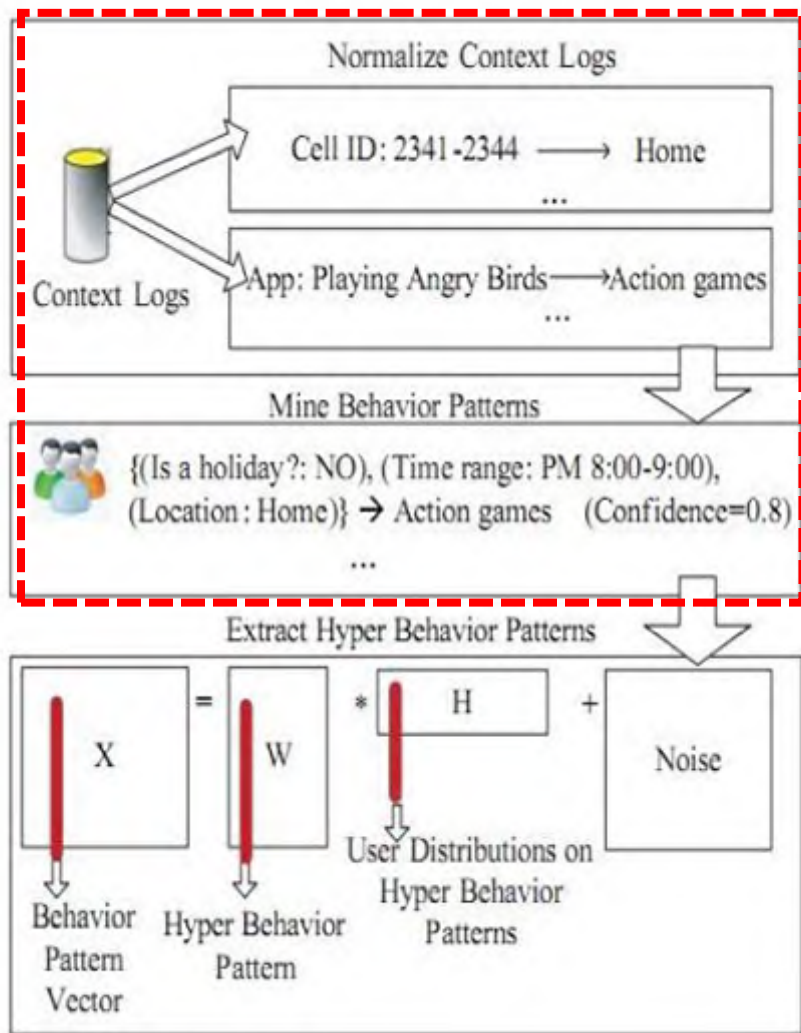
Data type	Contextual feature	Value range
Time Info	Day Name	{Mon., Tues., Wed., Thur., Fri., Sat., Sun.}
	Holiday	{Yes, No}
	Day Period	{Morning(AM7:00-AM11:00), Noon(AM11:00-PM14:00), Afternoon(PM14:00-PM18:00), Evening(PM18:00-PM21:00), Night(PM21:00-Next day AM7:00)}
	Time Range	{AM0:00-AM1:00, AM1:00-AM2:00, ..., PM23:00-PM24:00}
System Info	Profile Type	{General, Silent, Meeting, Outdoor, Pager, Offline}
	Battery Level	{Low(<35%), Middle(35%-65%), High(65%-85%), Full(>85%)}
	Inactive Time	{Short(< 10 minutes), Middle(10-30 minutes), Long(> 30 minutes)}
	Ring Type	{General, Ascending, Ring once, Beep, Silent}
	Ring Level	{Low(<35%), Middle(35%-75%), High(>75%)}
GSM Info	Cell ID	Integers.
	Area ID	Integers.
GPS Info	Speed	{Low(< 5km/h), Middle(5-20km/h), High(> 20km/h)}
	Movement	{Yes, No}
Event	Coordinate	Pair of longitude and latitude.
	Applications	{Call, Message, Game, Web browser, Music, Camera, etc.}

移动相似用户挖掘挑战：

1. 情境数据来源多：**多源异构**
2. 单个用户数据量少：**数据稀疏**
3. 移动用户的增多：**数据量大**

相似行为习惯的移动用户挖掘

带约束条件的贝叶斯矩阵分解：



1. 规范化情境日志

a). 位置情境规范化

用户位置标示映射到更高具有抽象意义的位置空间

b). 交互记录规范化

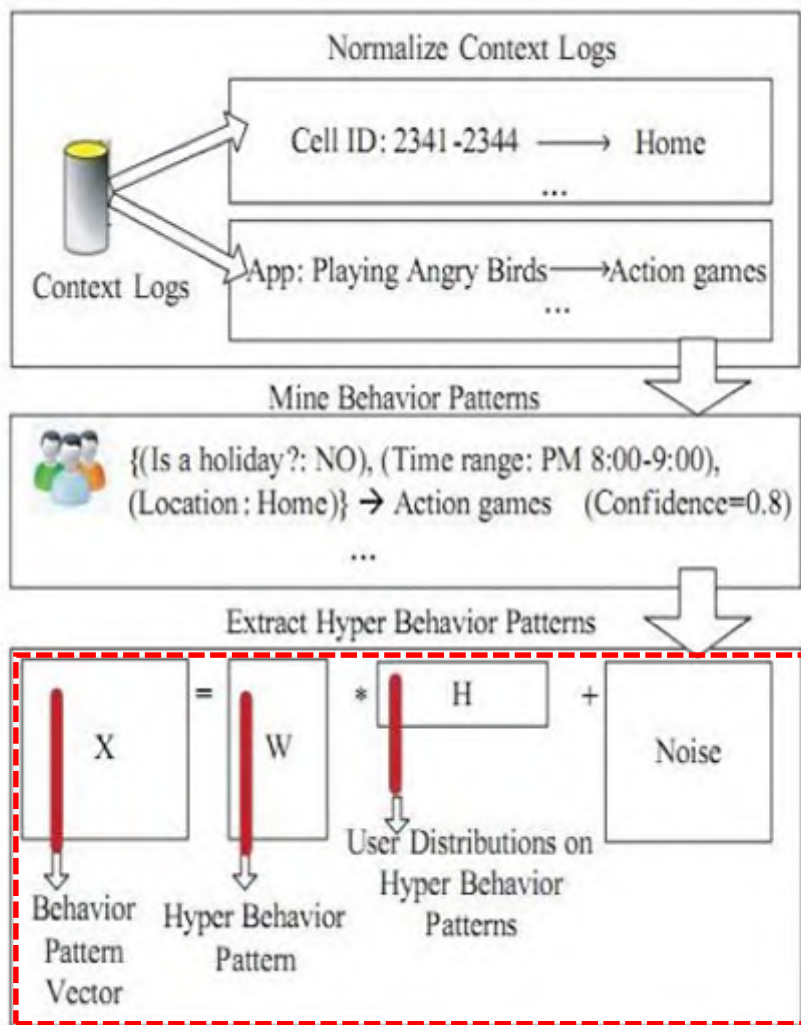
将用户交互记录映射到更高抽象级别的交互类别

c). 规范化情境日志的好处

减少数据的**稀疏性**，数据抽象具有更好**解释性**，提高数据的**泛化性能**

相似行为习惯的移动用户挖掘

带约束条件的贝叶斯矩阵分解：



1. 规范化情境日志

a). 位置情境规范化

用户位置标示映射到更高具有抽象意义的位置空间

b). 交互记录规范化

将用户交互记录映射到更高抽象级别的交互类别

c). 规范化情境日志的好处

减少数据的**稀疏性**，数据抽象具有**更好解释性**，提高数据的**泛化性能**

2. 带约束条件的贝叶斯矩阵分解

$$\Omega_{N \times m} = \Phi_{N \times K} \times \Theta_{K \times M} + N_{N \times M}$$

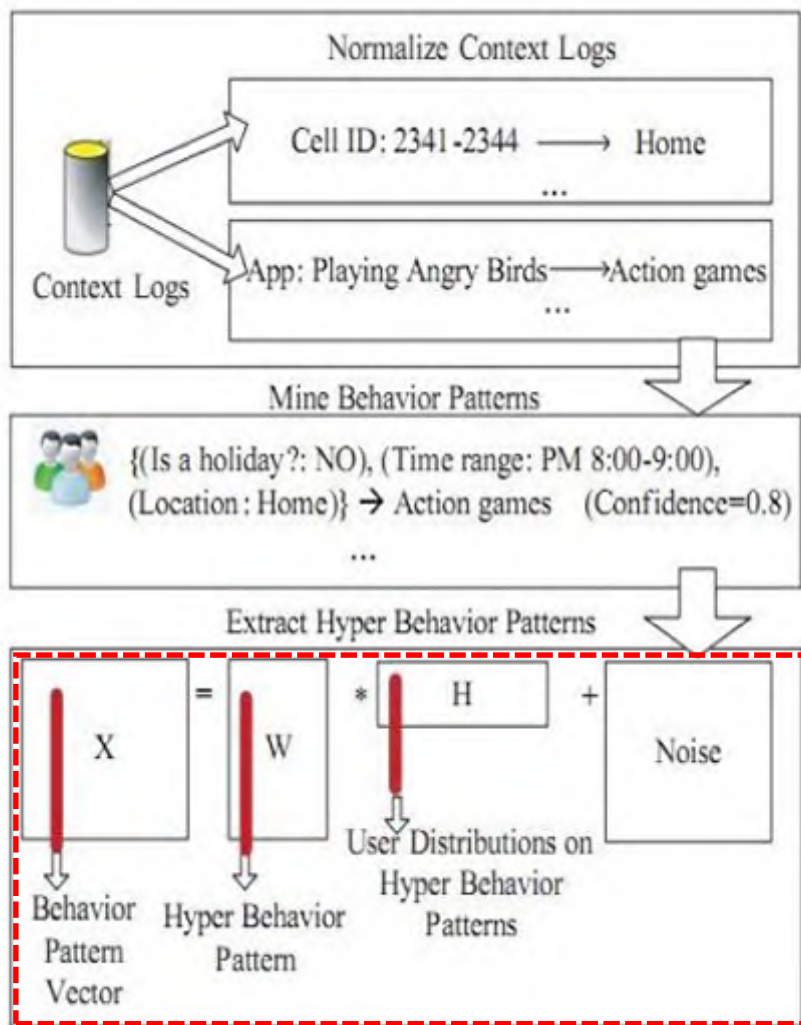
$\Omega_{N \times M}$ ：用户-行为模式矩阵

$\Phi_{N \times K}$ ：用户共同行为模式类矩阵

$\Theta_{K \times M}$ ：用户在共同行为模式上的概率分布矩阵

相似行为习惯的移动用户挖掘

带约束条件的贝叶斯矩阵分解：



1. 规范化情境日志

a). 位置情境规范化

用户位置标示映射到更高具有抽象意义的位置空间

b). 交互记录规范化

将用户交互记录映射到更高抽象级别的交互类别

c). 规范化情境日志的好处

减少数据的**稀疏性**，数据抽象具有**更好解释性**，提高数据的**泛化性能**

2. 带约束条件的贝叶斯矩阵分解

$$\Omega_{N \times m} = \Phi_{N \times K} \times \Theta_{K \times M} + N_{N \times M}$$

$\Omega_{N \times M}$ ：用户-行为模式矩阵

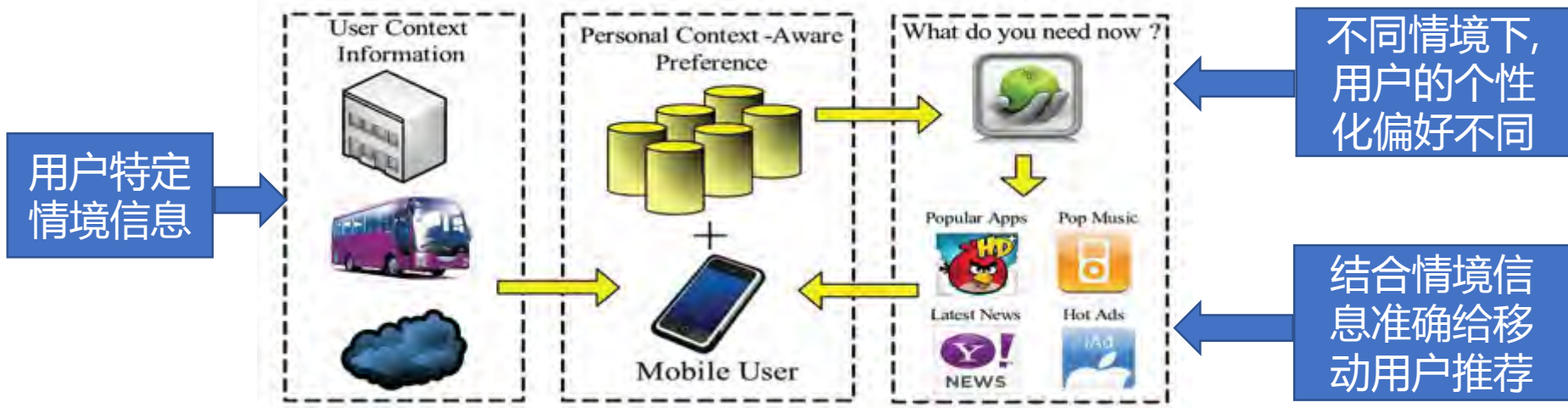
$\Phi_{N \times K}$ ：用户共同行为模式类矩阵

$\Theta_{K \times M}$ ：用户在共同行为模式上的概率分布矩阵

根据用户在共同行为模式上的概率分布得到相似用户

移动用户的个性化偏好挖掘

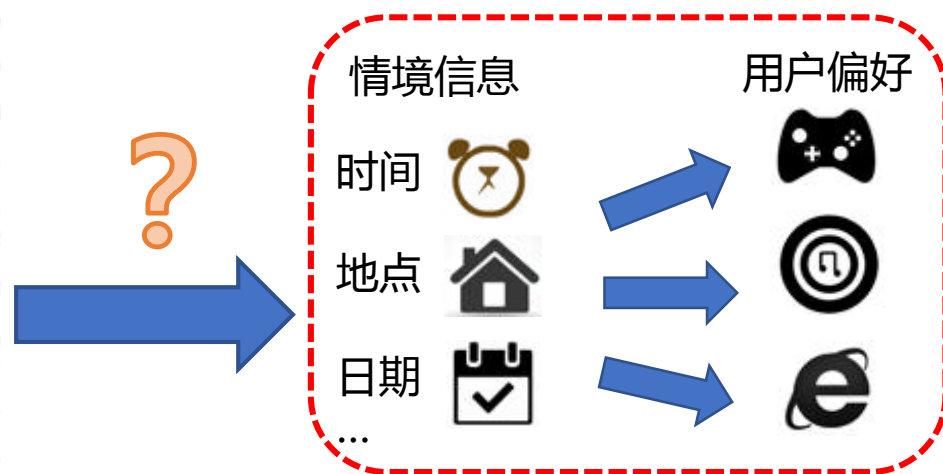
基于情境的移动用户个性化推荐



如何挖掘移动用户的个性化偏好？

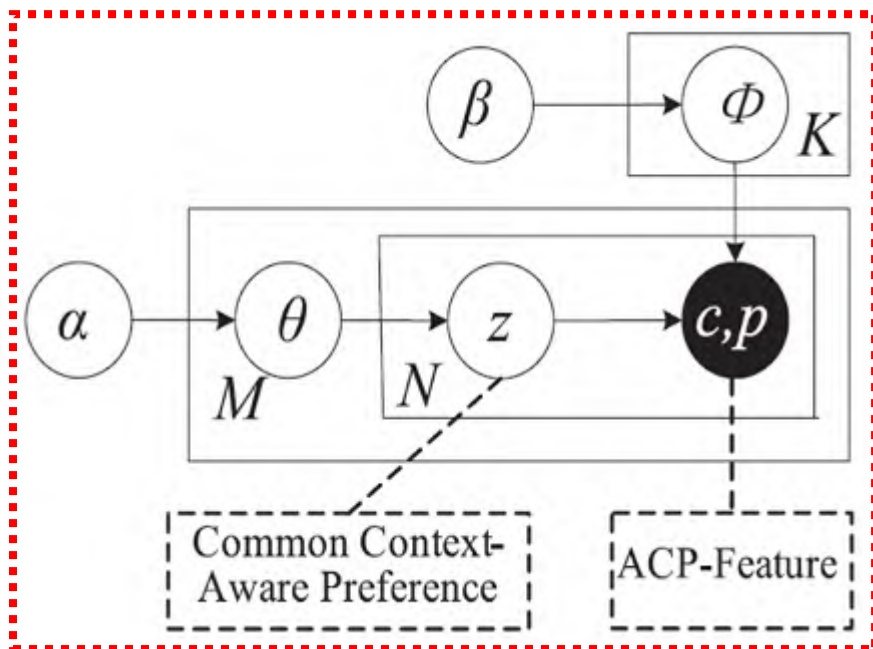
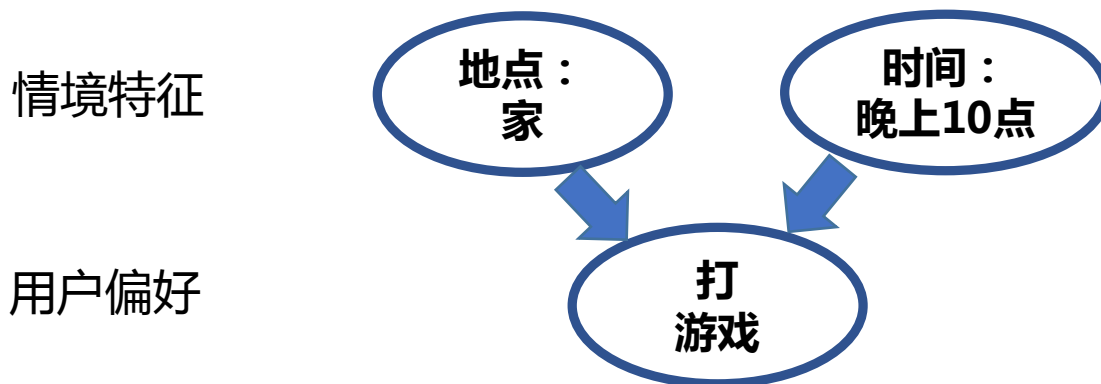
Timestamp	Context	Activity Records
t_1	{(Day name: Monday), (Time range: AM8:00-9:00), (Profile: General), (Battery: 5), (Location: Home)}	Null
t_2	{(Day name: Monday), (Time range: AM8:00-9:00), (Profile: General), (Battery: 5), (Location: On the way)}	Play action games (Fruit Ninja)
t_3	{(Day name: Monday), (Time range: AM8:00-9:00), (Profile: General), (Battery: 5), (Location: On the way)}	Null
t_{300}	{(Day name: Monday), (Time range: AM10:00-11:00), (Profile: Meeting), (Battery: 4), (Location: Meeting room)}	Null
t_{360}	{(Day name: Monday), (Time range: AM10:00-11:00), (Profile: Meeting), (Battery: 4), (Location: Meeting room)}	Visit websites
t_{445}	{(Day name: Monday), (Time range: AM11:00-12:00), (Profile: General), (Battery: 4), (Location: Work place)}	Play with SNS (Facebook)
t_{449}	{(Day name: Monday), (Time range: AM11:00-12:00), (Profile: General), (Battery: 4), (Location: Work place)}	Null

用户情境记录



移动用户的个性化偏好挖掘

方法一：情境特征相互独立假设



概率主题模型建模

$$P(a|C, u) \propto \sum_z \left(\prod_{p_i \in C} P(a, p_i|z) \cdot P(z|u) \right)$$

$P(a|C, u)$: 用户 u 在情境 C 下对活动 a 的偏好

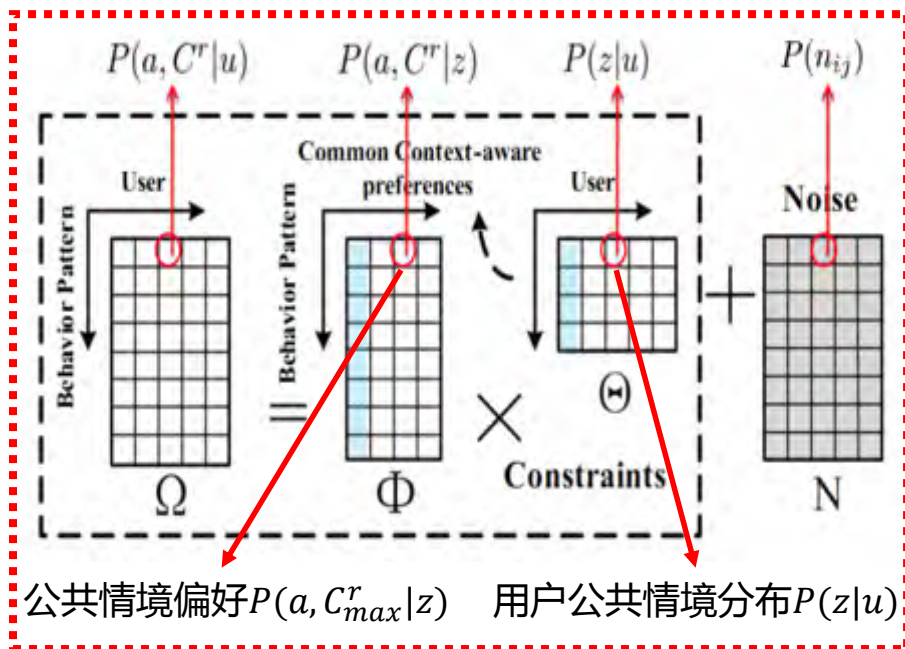
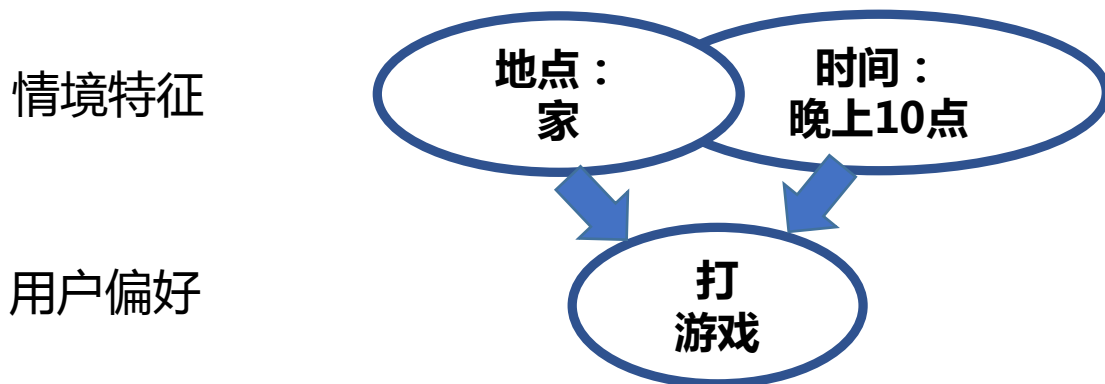
$P(a, p_i|z)$: 公共情境 z 上情境 p_i 和活动 a 分布

$P(z|u)$: 用户 u 的个性化偏好分布

根据用户在特定情境下对活动的偏好通过概率主题模型挖掘到用户的个性化偏好

移动用户的个性化偏好挖掘

方法二：情境特征相互影响假设



带约束的矩阵分解建模

$$\Omega_{N \times M} = \Phi_{N \times K} \times \Theta_{K \times M} + N_{N \times M}$$

$\Omega_{N \times M}$ ：用户 u 行为模式分布

$\Phi_{N \times K}$ ：公共模式上不同行为模式的分布

$\Theta_{K \times M}$ ：用户 u 的个性化偏好分布

$N_{N \times M}$ ：噪声分布

根据用户行为模式分布通过带约束的矩阵分解模型挖掘到用户的个性化偏好

移动商务大数据应用实例

■ 示范应用：面向移动App的排名欺诈检测

- 为了在移动App商店的应用排行榜上吸引更多用户，许多厂商雇用商业公司，通过“机器人”或者“水军”来刷榜，实现移动App商店排名恶意欺诈
- 如何检测这些排名欺诈，具有很多挑战：
 - 刷榜行为并不是长时间的，而是在某些时间段实现
 - 移动App数量众多，难以手动标注刷榜行为，需要设计自动的增量式方法实现
 - 移动App排行具有高度的动态性，需要寻找刷榜行为证据



1

大数据时代背景及现状

2

基于认知的教学个性化推荐

3

情景感知的移动用户推荐

4

金融领域带风险约束的推荐

5

结合社交因素的用户推荐

6

总结与展望

金融大数据应用与推荐系统的关系

■ 金融领域的应用与推荐系统的结合十分紧密

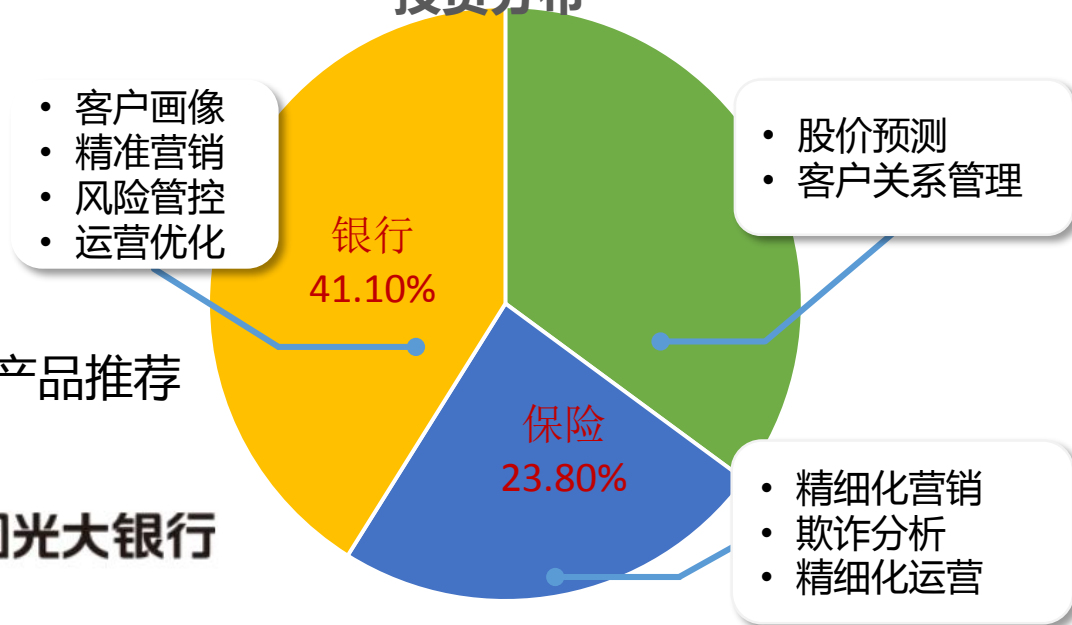
- 主要集中在银行、证券和保险三个部分
- 其中，精准营销、精细化运营、运营优化等都涉及到推荐系统

■ 应用案例:

- 中信银行进行实时的产品推荐，带来了更大的收益
- 光大银行建立了社交网络信息数据库，更有效地进行了产品推荐

中国金融行业大数据应用

投资分布



数据来源:赛迪顾问, 2012年

如何结合金融领域特点进行推荐

■ 在金融领域进行推荐的特点

- 投资人的特点：不同用户有不同的**风险偏好**、**预期收益**
- 投资产品的特点：产品本身具有**时效性**、**时变性**

挑战：如何结合用户特点进行产品的推荐，进一步进行精准营销？
如何结合产品特点进行客户的挖掘推荐，实现效益最大化？

■ 结合金融领域特点在**借贷投资方面**进行推荐的**两个场景**

- 满足投资人的**风险偏好**、**预期收益**的情况下进行投资产品**个性化推荐**
- 在进行营销前，在产品的销售期内进行**潜在投资人的挖掘推荐**



场景1：投资产品个性化推荐



场景2：投资人挖掘推荐

场景1：满足客户偏好进行金融产品个性化推荐

不同投资产品存在的**风险**不同，而投资人对风险的**偏好**也不同

Rate	Term	FICO®	Amount	Purpose	% Paid	Amount	Time Left
B 2 9.16%	60	725-729	\$28,000	Loan Refinancing & Consolidation	90%	\$2,575	6 days
B 2 11.47%	60	718-714	\$26,500	Loan Refinancing & Consolidation	82%	\$3,625	9 days
A 5 7.89%	60	795-799	\$16,800	Loan Refinancing & Consolidation	65%	\$5,825	9 days
A 5 7.89%	60	775-779	\$21,000	Other	95%	\$7,150	9 days
A 4 7.99%	60	715-719	\$14,000	Loan Refinancing & Consolidation	55%	\$6,300	10 days
A 4 7.89%	36	735-739	\$8,400	Loan Refinancing & Consolidation	94%	\$450	13 days
A 4 7.99%	60	725-729	\$28,800	Loan Refinancing & Consolidation	39%	\$17,500	9 days
B 1 8.99%	60	695-699	\$26,000	Medical Expenses	62%	\$9,875	9 days
A 3 7.59%	36	795-799	\$5,800	Loan Refinancing & Consolidation	89%	\$625	12 days
A 4 7.99%	60	795-799	\$10,000	Major Purchase	62%	\$3,775	10 days
A 5 7.89%	60	795-799	\$24,000	Loan Refinancing & Consolidation	59%	\$11,850	10 days
A 4 7.99%	60	725-729	\$25,000	Loan Refinancing & Consolidation	29%	\$18,750	10 days
A 5 7.89%	60	695-699	\$30,000	Home Improvement	22%	\$23,225	9 days
A 5 7.89%	60	680-684	\$16,000	Loan Refinancing & Consolidation	21%	\$12,500	13 days
A 5 7.89%	60	715-719	\$35,000	Credit Card Payoff	20%	\$27,750	9 days

Alice在投资时主要考虑

两个方面



哪些投资产品比较好？

如何组成最优组合？

由于自由的投资交易，对众多投资产品，不同投资人在选择时有不同的风险偏好

场景1：满足客户偏好进行金融产品个性化推荐

对投资产品的风险考虑

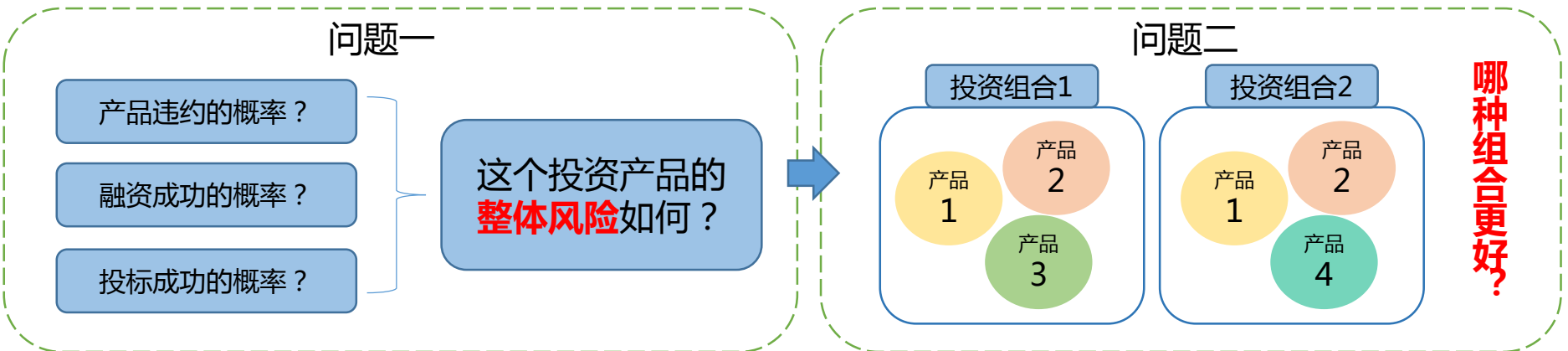
- 投资产品违约的风险
- 投资产品融资成功的风险
- 投资人投标成功的风险



产品会违约吗？

融资能成功吗？

投标了能成功吗？



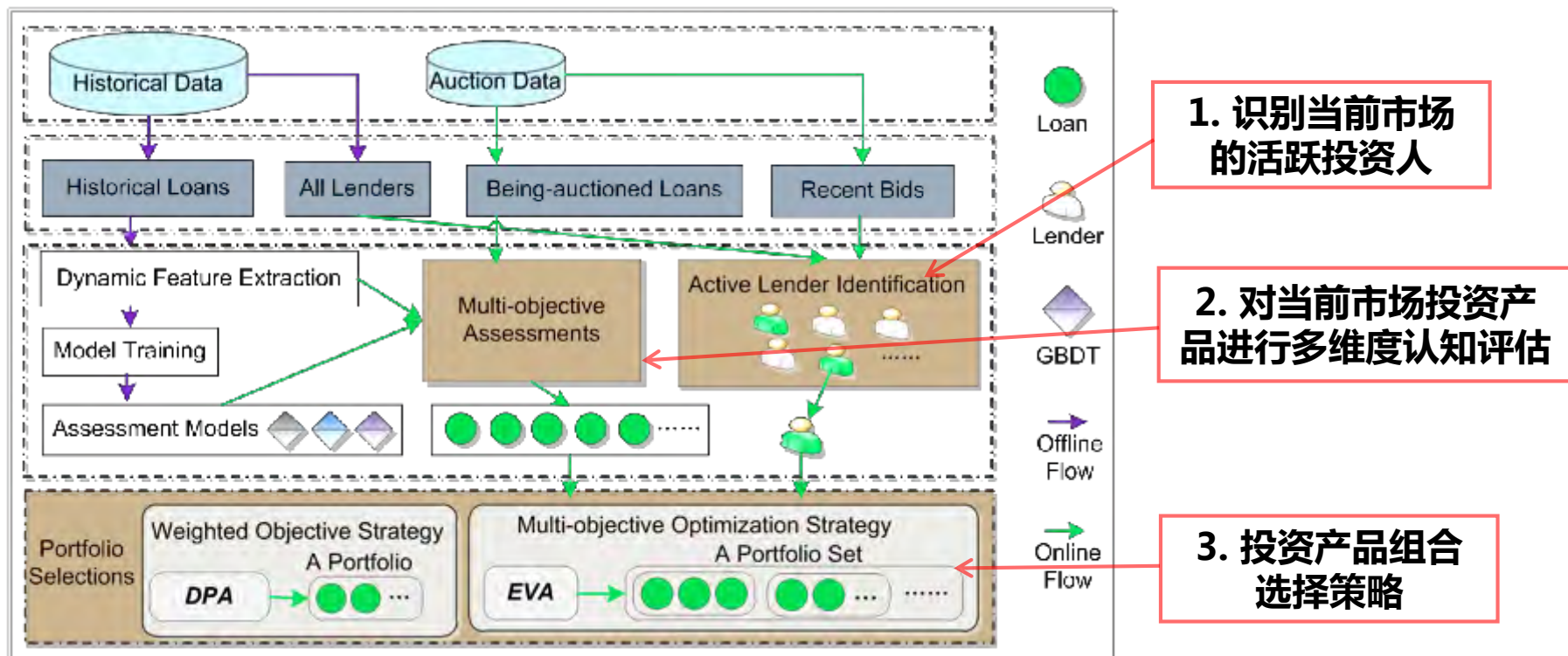
问题

- 如何评估投资产品在多个风险层面的表现
- 如何帮助投资人选择在她多方面风险考虑下最优的投资产品组合

场景1：满足客户偏好进行金融产品个性化推荐

解决方法（分三步）：

- 根据投资人最近的投资记录，识别当前市场的活跃投资人
- 针对问题一的解决方法：对当前市场的投资产品进行**多维度风险评估**（GBDT算法）
- 针对问题二的解决方法：**两种选择策略**（加权目标策略，多目标优化策略）



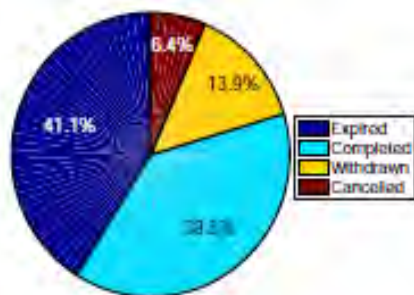
场景2：结合产品特性进行潜在投资人的推荐

什么是潜在投资人推荐？

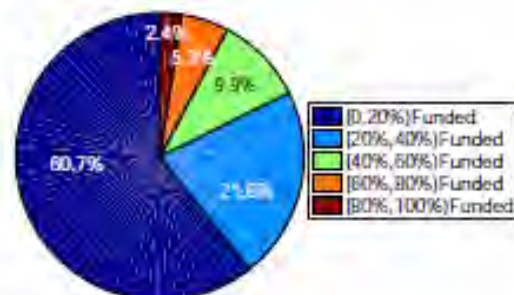
- 在众多投资人中，为投资产品找到潜在会投资该产品的投资人

为什么要进行潜在投资人推荐？

- 大多数投资产品，由于在投资期限内没有被潜在的投资人发现，导致投资失败，为投资产品推荐潜在的投资人可以[提高产品融资成功率](#)
- 有利于进行[精准营销](#)



41%的投资产品募集失败



投资失败的投资产品中有60%的产品只收到了20%以下的目标金额投资

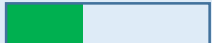


以借贷平台prosper为例

场景2：结合产品特性进行潜在投资人的推荐

问题

- 问题一：投资产品随着时间推移，投资的金额、投资的人数等会进行实时变化，考虑如何刻画投资产品的实时变化
- 问题二：投资产品投资时间有限，具有时效性
- 问题三：如何结合潜在投资人的特点，进行投资人推荐

投资过程动态变化

时间	产品融资比例	产品融资金额
10.8 9: 00	 33%	3300
10.8 18: 00	 50%	5000
10.9 12: 00	 82%	8200

时
变
性

投资时间有限

产品	融资时间	融资天数
产品1	2013.10.12 - 2013.12.12	60
产品2	2014.01.02 - 2014.02.01	30
产品3	2015.07.05 - 2015.09.05	60

时
效
性

结合以上投资产品特点和投资
人双方特点，推荐潜在投资人



场景2：结合产品特性进行潜在投资人的推荐

解决方法：混合随机游走方法

解决了投资产品动态变化的问题
(问题一)

动态推荐模型

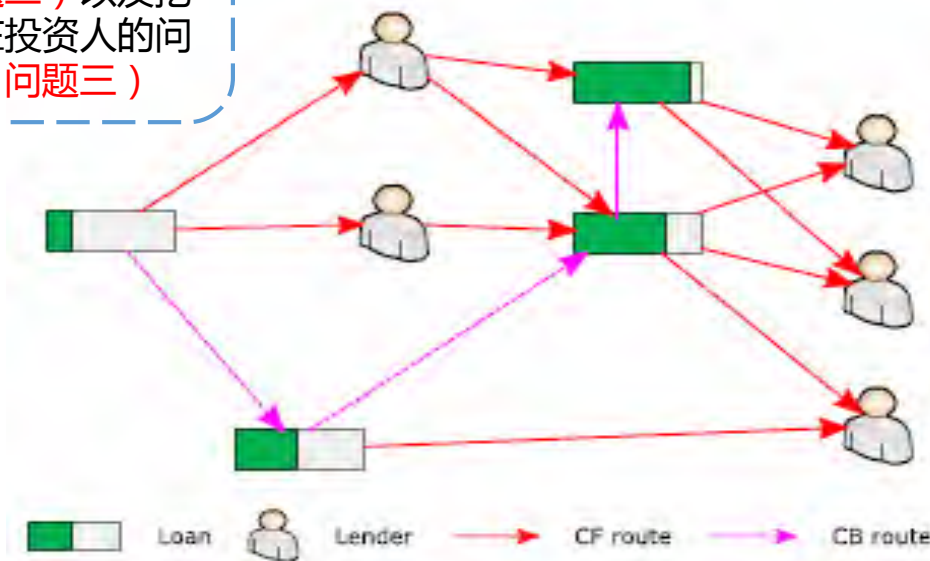
为投资产品建立实时的动态推荐模型

解决了投资产品募集时间有限的问题
(问题二) 以及挖掘潜在投资人的问题
(问题三)

混合随机游走模型

进行潜在投资人的推荐

- 结合协同过滤和基于内容的过滤方法
- 协同过滤的方法实现投资产品与投资人之间的游走
- 基于内容的过滤方法实现投资产品之间的游走

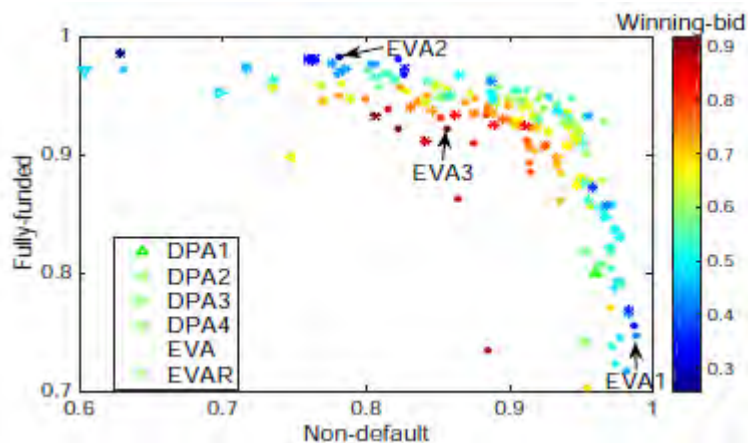


混合随机游走模型示意图

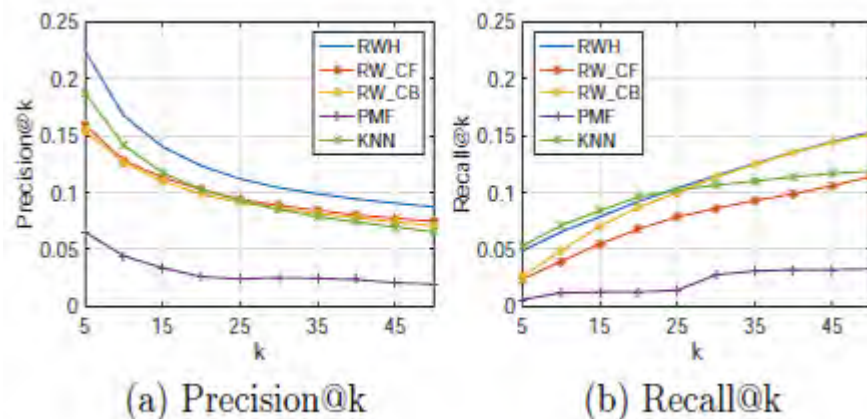
金融领域推荐应用场景结果

场景1的结果

- 在考虑三个风险的情况下，投资产品组合都落在了帕累托最优边界上
- 所提出的EVA方法结果更优，说明了该方法应用在投资产品组合推荐上的有效性



场景2的结果



- 比较所提出的混合随机游走算法 (RWH) 与其他算法在潜在投资人推荐的结果
- 在精度和召回率上，RWH都表现得优于其他算法

通过结合金融领域特点进行推荐，实现了投资人和投资产品双方效益的最大化

1

大数据时代背景及现状

2

基于认知的教学个性化推荐

3

情景感知的移动用户推荐

4

金融领域带风险约束的推荐

5

结合社交因素的用户推荐

6

总结与展望

大数据时代下的社交推荐

■ 大数据时代下社交网络与推荐系统紧密结合

- Facebook等社交网络大行其道，而其中包含了海量的数据
- 个性化推荐在社交网络上取得了广泛应用

应用案例:

- Facebook与亚马逊合作进行实时的产品推荐，带来了更大的收益
- 阿里巴巴建立了社交网络信息数据库，更有效地进行产品推荐

社交大数据



个性化
推荐

推荐系统



社交推荐的特点和挑战

什么是社交推荐：

- 结合用户之间的**社交关系**进行个性化推荐

社交推荐推荐什么：

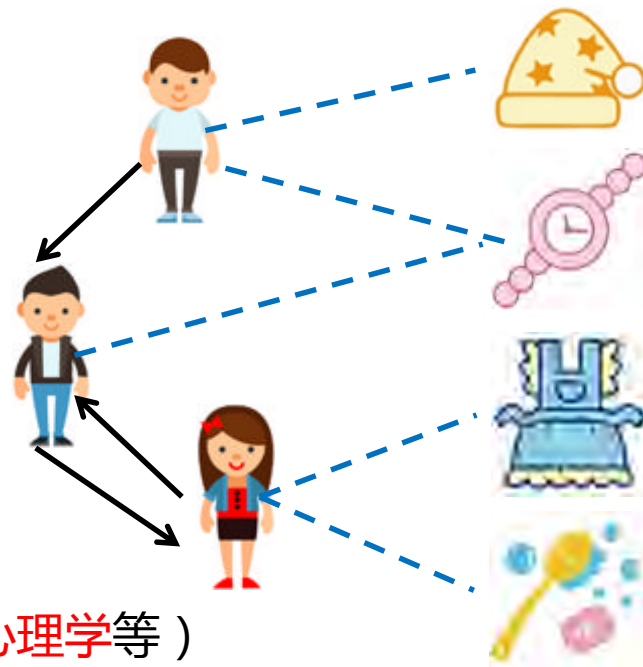
- 推荐与用户行为相关的事物
- 比如推荐好友，推荐商品

社交推荐的特点：

- 有特殊的领域知识（社交影响力传播、**社会心理学**等）
- 推荐结果的可解释性较好（如推荐理由是某个好友喜欢）

社交推荐的挑战：

- 如何刻画社交影响和同质性之间的相互作用
- 如何衡量社交网络**动态演化**过程及其影响
- 如何利用**社会心理学建模**用户行为与社交关系的共同演化



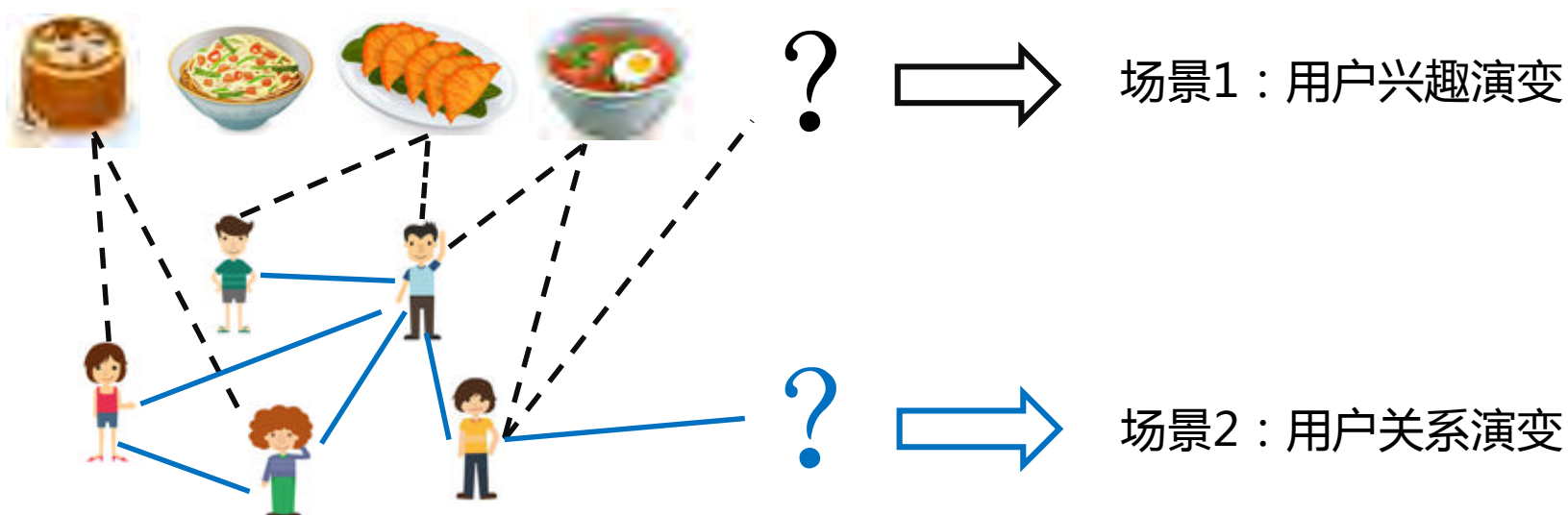
动态社交影响

传统方法的局限性和不足之处：

- 传统的社交推荐并没有很好地建模社交影响和用户行为演变之间的关系
- 传统的社交推荐忽视了用户受社交影响之后行为变化对社交关系的反作用

动态社交推荐的特点：

- 更真实地刻画社交网络与用户行为之间的相互演变过程
- 用户社交心理（如从众）的影响会与社交关系强弱的变化而联合演变
- 更好地解释用户行为演变的机制



场景1：用户兴趣演变预测

问题定义

- 在社交服务平台的时间轴 1 到 T 上，给定用户消费张量C(user-item-time)及社交张量S(user-user-time)
- 预测用户在 T + 1 时间的**消费行为**

基本思路

- 每个用户在时刻t有一个潜在的兴趣因素 U^t
- 兴趣因素以及社交因素预测用户消费行为

用户**消费行为**演化分析

- 用户更喜欢和历史消费兴趣相似的产品
- 用户行为受**社交好友**影响

$$p(U_a^t) = \mathcal{N}(U_a^t | \bar{U}_a^t, \sigma_U^2 \mathbf{I})$$

$$\text{where } \bar{U}_a^t = (1 - \alpha_a) U_a^{(t-1)} + \alpha_a \sum_{b \in N_a^{(t-1)}} F_{ab}^t U_b^{(t-1)}$$

协同过滤

社交影响

场景2：用户关系演变预测

问题定义

- 在社交服务平台的时间轴 1 到 T 上，给定用户消费张量 C(user-item-time) 及社交张量 S(user-user-time)
- 预测用户在 T + 1 时间的**社交链接行为**

基本思路

- 每个用户在时刻 t 有一个潜在的兴趣因素 U^t
- 邻近用户以及消费行为预测用户社交行为

用户**社交行为**演化分析

- 用户喜欢和拥有**相同兴趣**的顶点建立链接
- 用户喜欢和在社交网络中距离更近的顶点建立链接

$$\hat{S}_{ab}^t = (1 - \beta_a)h(a, b, t) + \beta_a \langle U_a^{(t-1)}, U_b^{(t-1)} \rangle$$

任意节点邻近模型

同质性

动态演化

相互作用动态演化

- 该时刻的消费行为影响下一个时刻的社交关系
- 该时刻的社交关系影响下一时刻的消费行为



动态演化

消费行为演化

- 更喜欢历史兴趣相似的产品（协同过滤）
- 行为受社交邻居影响（社交影响）

链接行为演化

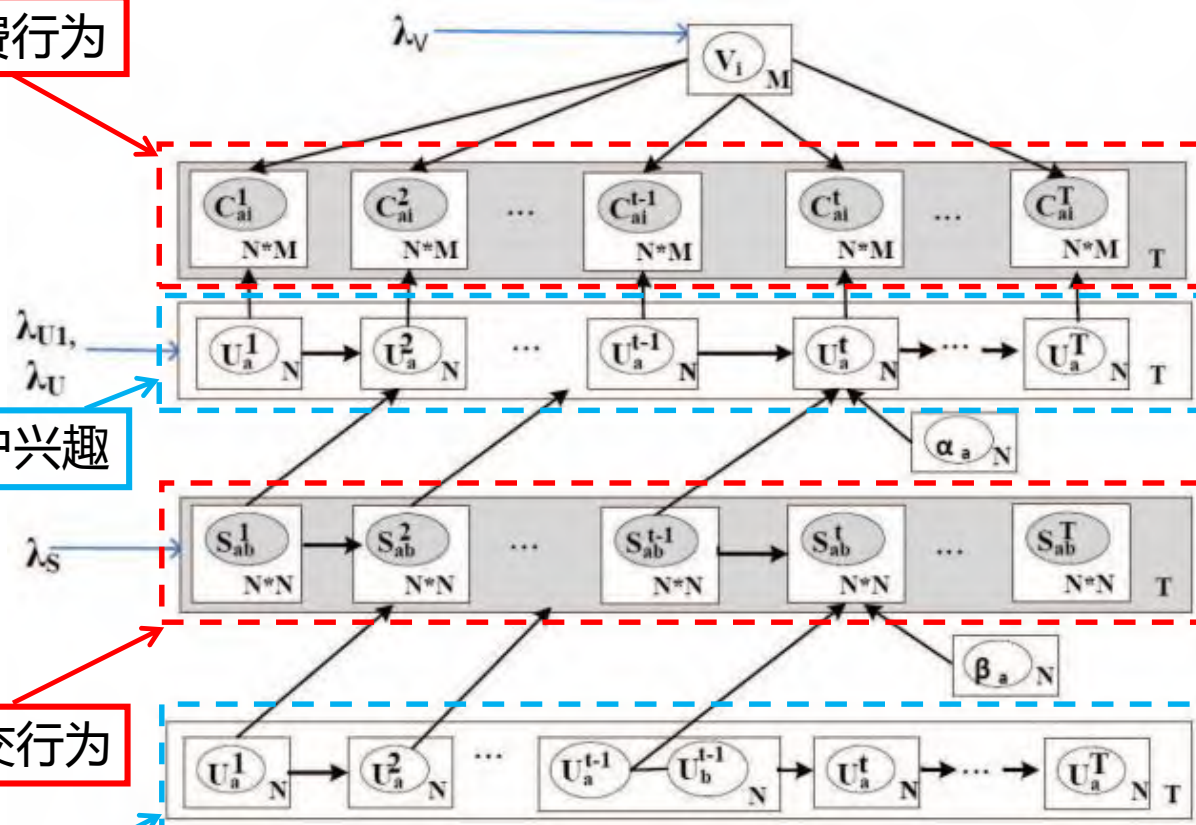
- 喜欢和社交距离近的用户（节点接近）
- 喜欢兴趣相投的用户（同质性）

消费行为

用户兴趣

社交行为

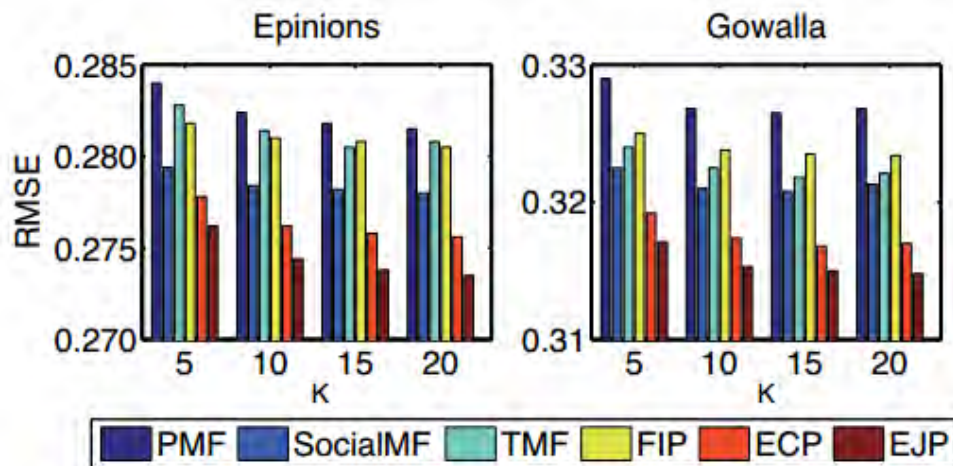
用户兴趣



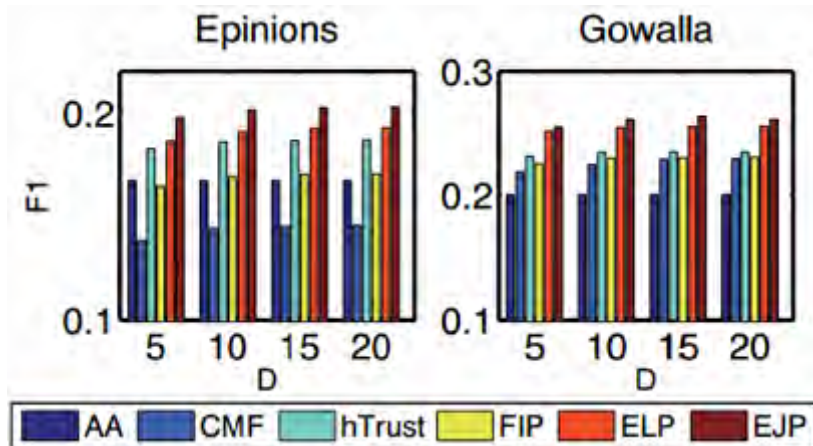
效果展示

场景1的结果

- 在RMSE指标下，商品推荐都在演化联合预测（EJP）算法上获得了**最优结果**
- 所提出的**EJP方法结果更优**，说明了该方法应用在商品推荐上的有效性



场景2的结果



- 比较所提出的演化联合预测算法（EJP）与其他算法在好友推荐的结果
- 在F1指标上，EJP都表现得优于其他算法

通过结合社交领域特点进行推荐，实现了用户兴趣及关系演变的精准预测

1

大数据时代背景及现状

2

基于认知的教学个性化推荐

3

情景感知的移动用户推荐

4

金融领域带风险约束的推荐

5

结合社交因素的用户推荐

6

总结与展望

总结与展望

面对不同领域独特的应用场景，泛化的模型结果难以满足用户的需求。
通过结合各行业独特的领域知识，构建适用于不同领域下的推荐系统。

提高可解释性

- 通过结合领域知识，得到的结果具有更强的可解释性，利于方法的推广和说明。

加强针对性

- 通过针对独特的应用场景进行建模，可以解决更多泛化模型难以解决的问题。

有效数据筛选

- 通过将应用情景限制在一个特定的场景里，可以筛选掉大量的无关数据。



Thanks!