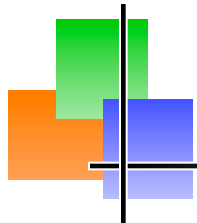


# 基于表示学习的信息推荐算法及应用

高升

北京邮电大学模式识别实验室

2016年11月



# 提纲



一. 研究背景

二. 信息推荐算法的表示学习模型

A. 基于矩阵分解的Aspect-level推荐算法

B. 基于偏好向量学习的信息推荐算法

三. 总结

# 研究背景 - 个性化推荐场景



在线试读

**创新之路:提升商业价值的八大创新法则** 平装 - 2015年12月18日  
毛正行 (MADGE M.MEYER) (作者), 侯伟鹏 (译者)  
分享我的评价 | 分享

显示所有 2 格式和版本

Kindle电子书 ¥30.52	平装 ¥35.90
---------------------	--------------

使用我们的 免费Kindle阅读软件

促销信息: 16万图书满100元减20元 共3个促销

配送至: 北京东城区: 现在有货

送达日期: 明天(11月11日), 请在3小时53分钟内下单并选择“快速送货上门”。  
(精确送达时间请于结账页面查询)

销售配送: 由亚马逊直接销售和发货。

← 当前兴趣

购买此商品的顾客也同时购买

根据当前兴趣进行推荐

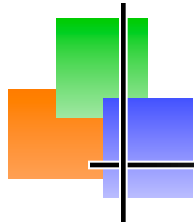
推荐的商品

重新定义创新  
谢德邦  
☆☆☆☆☆ 2  
精装  
¥42.00 Prime

一页纸企划书  
高桥宪行  
☆☆☆☆☆ 3  
平装  
¥33.10 Prime

重新定义工作: 大连接时代职业、公司和领导力的颠覆性变革  
【美】Jacob ...  
☆☆☆☆☆ 8  
平装  
¥38.60 Prime

狩猎式创新: 如何让你的创新思想源源不断  
杰里米·古奇 (J ...  
☆☆☆☆☆ 2  
平装  
¥29.30 Prime



# 研究背景 - 个性化推荐场景



您最近查看的商品和相关推荐

猜您喜欢

第 1 页, 共 2 页

根据浏览历史进行推荐



重新定义创新  
谢德芬  
★★★★☆ 2  
精装  
¥42.00 Prime



重新定义工作: 大连接时代  
职业、公司和领导力 ...  
【美】Jacob ...  
★★★★☆ 8  
平装  
¥38.60 Prime



一页纸企划书  
高桥宪行  
★★★★☆ 3  
平装  
¥33.10 Prime



狩猎式创新: 如何让你的创  
新思想源源不断  
杰里米·古奇 (J...  
★★★★☆ 2  
平装  
¥29.30 Prime



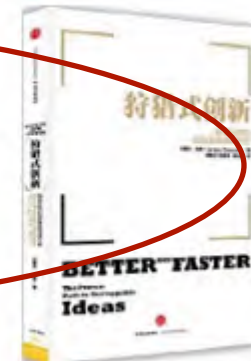
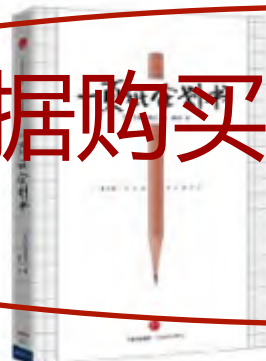
重新定义战略  
辛西娅·蒙哥马利  
★★★★☆ 3  
精装  
¥35.40 Prime



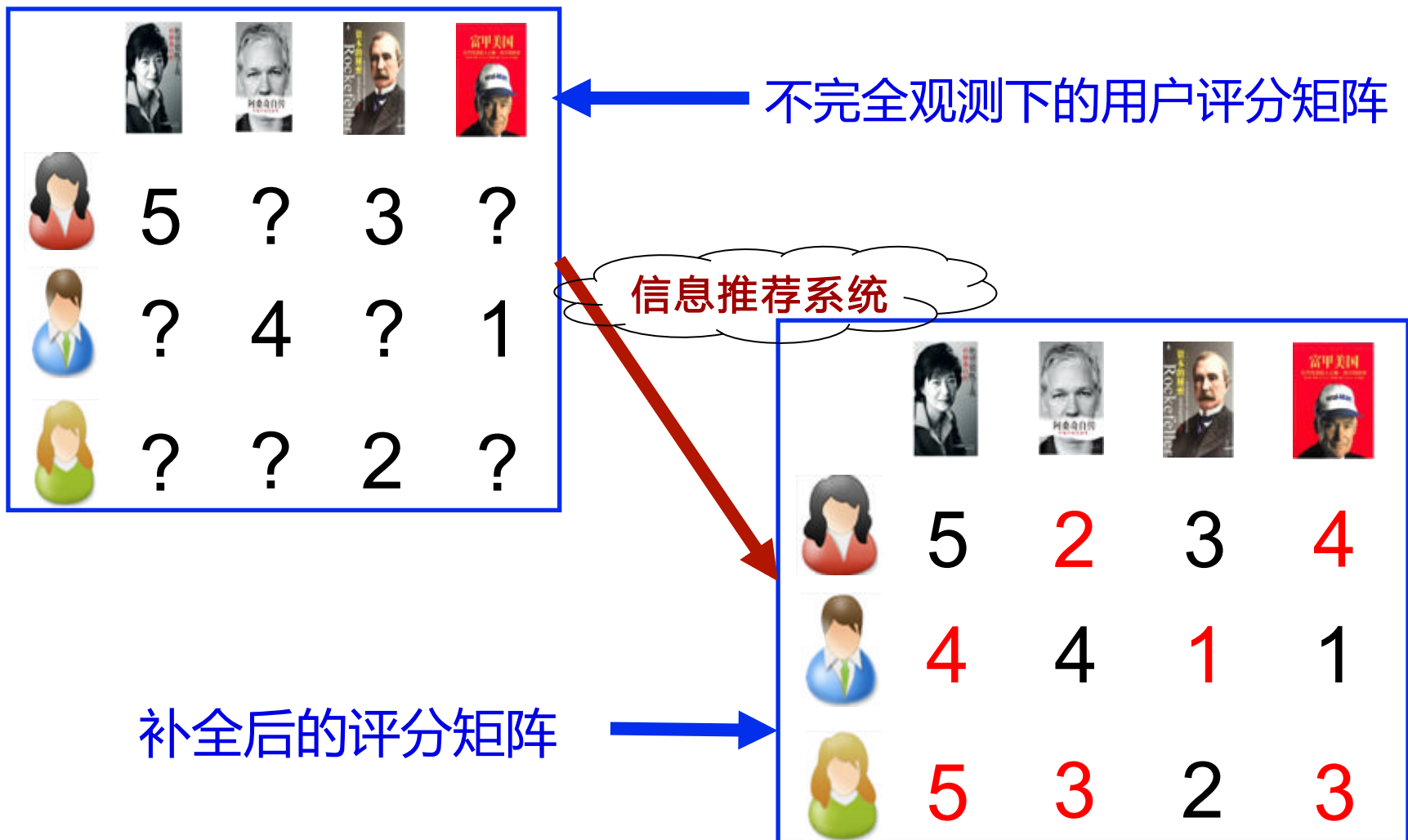
创新者系列(套装共3册)  
克莱顿·克里斯坦森 ...  
★★★★☆ 39  
精装  
¥106.00

购买 *创新之路:提升商业价值的八大创新法则* 的客户还购买了:

根据购买历史进行推荐

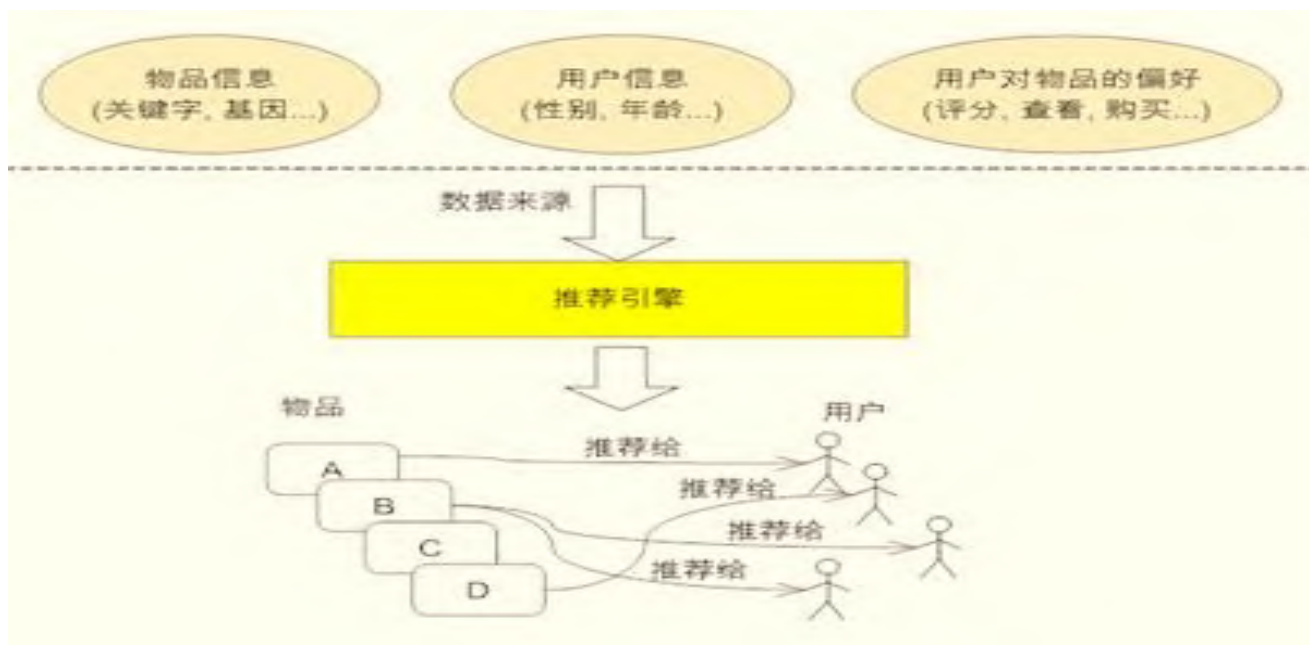


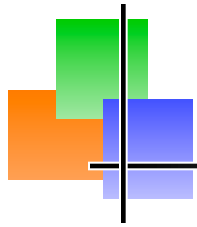
# 研究背景 - 个性化推荐任务



# 信息推荐系统

- 一般情况下，信息推荐系统所需要的数据源包括：
  - 用户对商品或者信息的偏好（用户行为信息）：根据应用本身的不同，包括用户对物品的评分、用户浏览物品的记录、用户的购买记录、用户的点击记录等。
  - 系统用户的基本信息：例如性别、年龄、职业、社交网络等。
  - 推荐对象的元数据信息：例如关键字，类别、属性描述等。





# 信息推荐系统



## □ 用户的偏好信息可以分为两类：

- **显式的用户反馈**：用户在网站上自然浏览或者使用网站以外，显式的提供反馈信息，例如用户对物品的评分、或者对物品的评论。
- **隐式的用户反馈**：用户在使用网站时产生的数据，隐式的反应了用户对物品的喜好，例如用户购买了某物品，用户浏览了某物品的网页、用户点击了某个商品等。

★★★★★ 1931条评论 人均：150元 口味：9.0 环境：9.1 服务：8.3

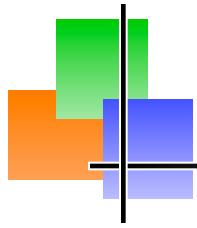
地址：海淀区 科学院南路2号融科资讯中心C座北楼1-2楼(近中科院...)

电话：010-82862800

aries喵 ★★★★★

★★★★★ 口味：4 环境：4 服务：4 人均：100

环境很安静，地方很干净，饭菜味道也很不错，非常喜欢的啦。  
和同事一起去的，定的位置，居然还有短信提醒，告诉我位子订好了。虽然现在禁烟，但是很多饭店还是可以抽烟，进入店.....



# 信息推荐系统



- 根据系统使用的数据源不同，信息推荐系统可以分为以下：
  - 根据系统用户的基本信息发现用户之间的相关程度，称为基于用户画像的信息推荐系统（User Profile-based Recommendation）
  - 根据推荐物品或内容的元数据，发现物品或者内容的相关性，称为基于内容的信息推荐系统（Content-based Recommendation）
  - 根据用户对物品或者信息的偏好，发现物品或者内容本身的相关性，或者是发现用户的相关性，称为基于协同过滤的信息推荐系统（Collaborative Filtering-based Recommendation）



# 研究背景 - 表示学习

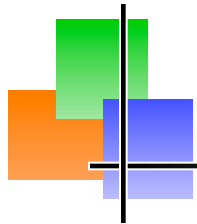


## □ 表示学习的目标

- 通过把目标对象映射到低维的统一特征空间，学习用户、商品以及用户偏好（行为）的潜在语义信息，从而进行相关性分析。

## □ 基本假设

- 把个性化信息推荐任务转为一个语义匹配的过程。
- 通过表示学习，将用户和商品转换为一个低维特征空间中的数值向量（分布式表示），那么基于用户偏好和商品属性的个性化推荐问题可以转换为语义特征向量关系的计算过程。



# 提纲



一. 研究背景

二. 个性化推荐的表示学习模型

A. 基于矩阵分解的Aspect-level信息推荐算法

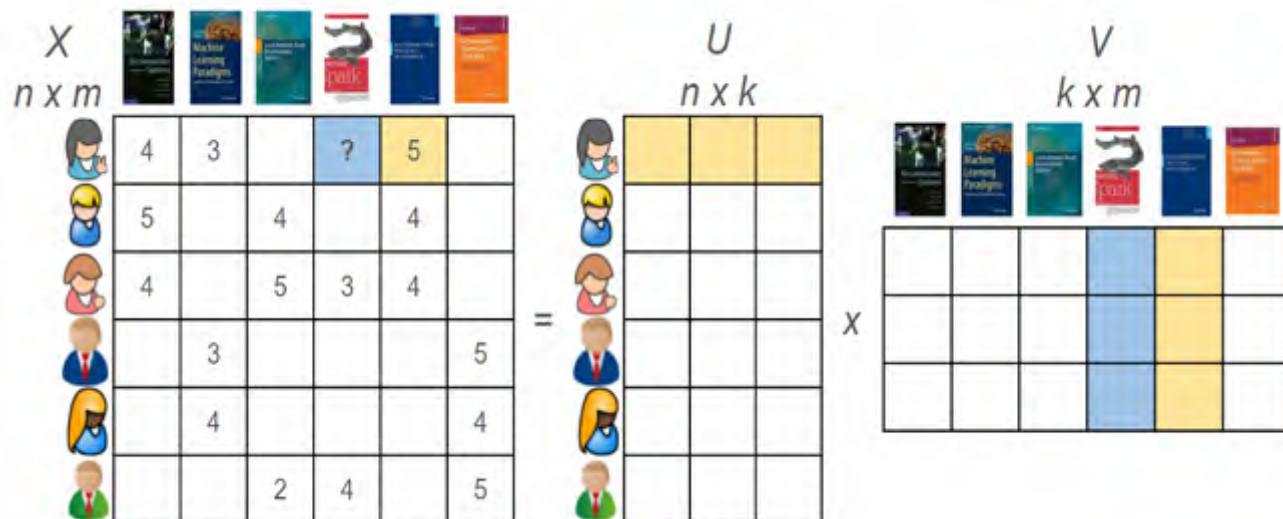
B. 基于偏好向量学习的个性化推荐算法

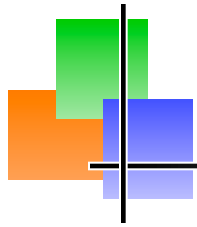
三. 总结

# 表示学习示例

## 基于矩阵分解的表示学习模型

- 矩阵分解（如奇异值分解，非负矩阵分解等）方法：将用户和商品都表示成统一潜在特征空间中的因子向量，这些潜在表示因子代表了用户如何对商品进行评分。评分的高低代表了用户对商品的倾向，由用户特征表示向量和商品特征表示向量的乘积确定。
- 给定用户和商品的潜在表示向量，通过向量间的内积运算可以预测用户将会给未评价的商品如何评分（用户的偏好或行为）。





# 信息推荐系统



## □ 用户的偏好信息可以分为两类：

- **显式的用户反馈**：用户在网站上自然浏览或者使用网站以外，显式的提供反馈信息，例如用户对物品的评分、或者对物品的评论。
- **隐式的用户反馈**：用户在使用网站时产生的数据，隐式的反应了用户对物品的喜好，例如用户购买了某物品，用户浏览了某物品的网页、用户点击了某个商品等。

★★★★★ 1931条评论 人均：150元 口味：9.0 环境：9.1 服务：8.3

地址：海淀区 科学院南路2号融科资讯中心C座北楼1-2楼(近中科院...)

电话：010-82862800

aries喵 ★★★★★

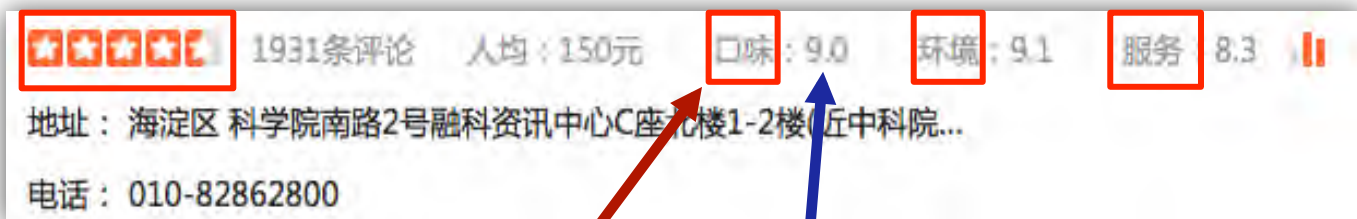
★★★★★ 口味：4 环境：4 服务：4 人均：100

环境很安静，地方很干净，饭菜味道也很不错，非常喜欢的啦。  
和同事一起去的，定的位置，居然还有短信提醒，告诉我位子订好了。虽然现在禁烟，但是很多饭店还是可以抽烟，进入店.....

# 如何利用评论中的语义信息？



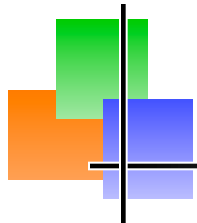
- 用户评论中蕴含了大量的语义信息，包括要素（Aspect）和情感（Sentiment）。



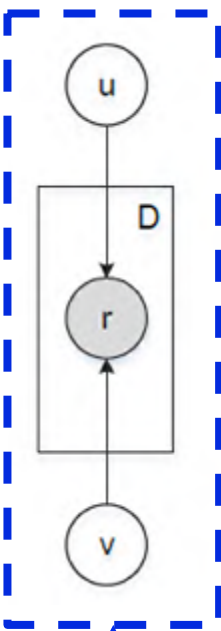
要素  
Aspect

情感  
Sentiment

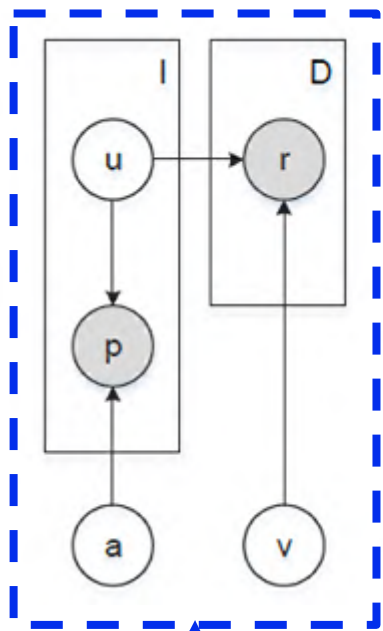
- ◇ 用户 — Aspect
- ◇ 商品 — Aspect



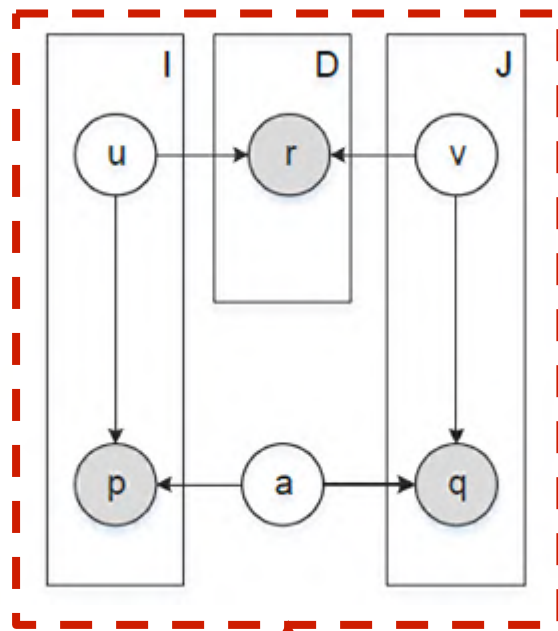
# 结合评论的信息推荐算法模型演化



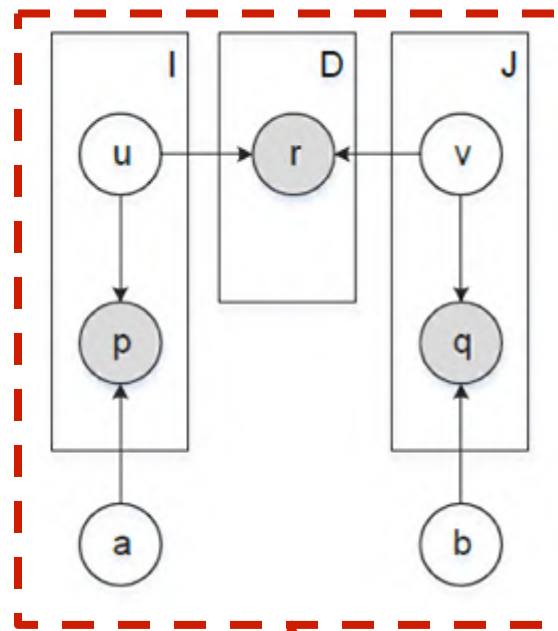
LFM



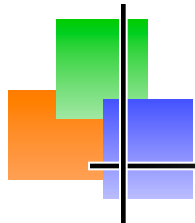
HFT



ALFM-P



ALFM-Q

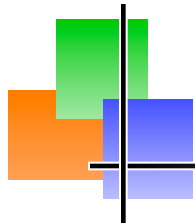


## □ 基于Aspect和矩阵分解的表示学习

$$\begin{aligned} \min \mathcal{J} = & \frac{1}{2} \left\{ \sum_{i=1}^I \sum_{j=1}^J c_{ij} (u_i v_j^T + \mu + \beta_i + \beta_j - r_{ij})^2 \right. \\ & + \alpha_u \sum_{i=1}^I \sum_{m=1}^M (u_i a_m^T - p_{im})^2 \\ & + \alpha_v \sum_{j=1}^J \sum_{m=1}^M (v_j a_m^T - q_{jm})^2 \\ & + \lambda_u \sum_{i=1}^I \|u_i\|^2 + \lambda_v \sum_{j=1}^J \|v_j\|^2 \\ & + \lambda_w \sum_{m=1}^M \|a_m\|^2 \\ & \left. + \lambda_b \left( \sum_{i=1}^I \beta_i^2 + \sum_{j=1}^J \beta_j^2 \right) \right\} \end{aligned}$$

矩阵分解

要素约束



# 基于矩阵分解的Aspect-level推荐算法



## □ 自动抽取评论中的Aspect和每个Aspect的对应评分

比邻故宫，地理位置相当优越...店家的景色也是得天独厚的...这是我等位最长的一次晚餐，整整两个半小时...点单超级迅速...北京炸酱面是到北京不得不吃的一道菜，店家做得很好吃...烤鸭不愧是店招的菜品，太好吃了...

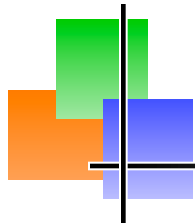
地理位置 正倾向 (4.8)      人流量 负倾向 (0.9)  
环 境 正倾向 (4.2)      服 务 正倾向 (4.6)  
口 味 正倾向 (5)

★★★★☆ 1931条评论 人均：150元 口味：9.0 环境：9.1 服务：8.3

地址：海淀区科学院南路2号融科资讯中心C座北楼1-2楼(近中科院...)

电话：010-82862800







# 基于矩阵分解的Aspect-level推荐算法



- 分析用户对Aspect的偏好，总结商品Aspect的评分。



环境：3.6      服务：1.7      口味：4.9

四季民福烤鸭店(故宫店)   手机买单 积分抵现

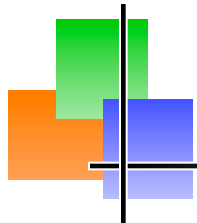
★★★★★ 4546条评论    人均：135元    口味：9.2    环境：9.4    服务：9.2

地址：南池子大街11号故宫东门旁

电话：010-65267369

环境：9.4      服务：9.2      口味：9.2

Qiu L, Gao S, Cheng W, et al. Aspect-based latent factor model by integrating ratings and reviews for recommender system[J]. Knowledge-Based Systems, 2016.



# 基于矩阵分解的Aspect-level推荐算法



- 计算用户的个性化评分，以及每个Aspect 的对应评分

aries喵 口味: 4 环境: 4 服务: 4 人均: 100

环境很安静，地方很干净，饭菜味道也很不错，非常喜欢的啦。  
和同事一起去的，定的位置，居然还有短信提醒，告诉我位子订好了。虽然现在禁烟，但是很多饭店还是可以抽烟，进入店.....

用户对每个Aspect的得分

## 总体评分

- 在获取到每个用户对Aspects的评分之后，可以得到Aspects的平均得分。
- 可以获取用户对商品的总体评分

# 基于矩阵分解的Aspect-level推荐算法

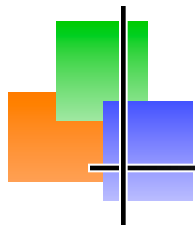


## □ 个性化推荐列表

The screenshot displays a list of restaurant recommendations. Each item includes a photo, name, location, star rating (highlighted with a red box), number of reviews, average price, and aspect-based scores for taste, environment, and service. The first item is '胡同小馆(东大桥店)' with a 4.5-star rating. The second is '四季民福烤鸭店(故宫店)' with a 4.9-star rating. The third is '吾家印象院派味道(健德门店)' with a 4.4-star rating.

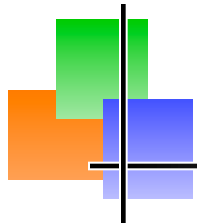
Restaurant Name	Star Rating	Reviews	Avg Price	Taste	Environment	Service
胡同小馆(东大桥店)	4.5	15	¥76	6.8	6.7	6.6
四季民福烤鸭店(故宫店)	4.9	5306	¥135	9.2	9.4	9.2
吾家印象院派味道(健德门店)	4.4	42	¥71	8.0	8.8	8.3

- 摒除了单纯以总评分进行推荐排序的方式。
- 考虑不同用户的不同需求，可以解释用户的偏好分布。
- 根据用户对不同Aspect的偏好，可以对待推荐商品进行排序。



## ■ 基于Aspect的景点推荐系统 Demo





# 提纲



一. 研究背景

二. 个性化推荐的表示学习模型

A. 基于矩阵分解的Aspect-level推荐算法

**B. 基于偏好向量学习的信息推荐算法**

三. 总结

# 信息推荐系统

## □ 多样化的用户偏好信息表示：

- 显式的用户反馈：用户在网站上自然浏览或者使用网站以外，显式的提供反馈信息，例如用户对物品的评分、或者对物品的评论。
- 隐式的用户反馈：用户在使用网站时产生的数据，隐式的反应了用户对物品的喜好，例如用户购买了某物品，用户浏览了某物品的网页、用户点击了某个商品等。



**I Like To Make Stuff**

@iliketomakestuff

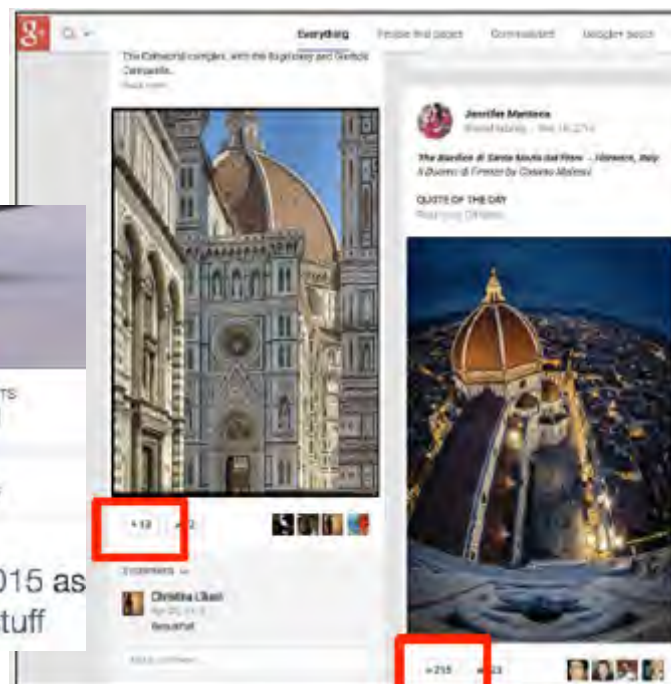
Maker/Dad/Musician/Runner/etc/etc/etc I make free how-to content

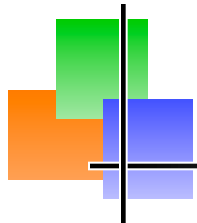
Savannah, GA

TWEETS	FOLLOWING	FOLLOWERS	FAVORITES	LISTS
2,876	182	2,701	1,686	1

Tweets Tweets & replies Photos & videos

I Like To Make Stuff @iliketomakestuff · 2h  
If you see me at #makerfaire2015 as a sticker!! #iltms #iliketomakestuff

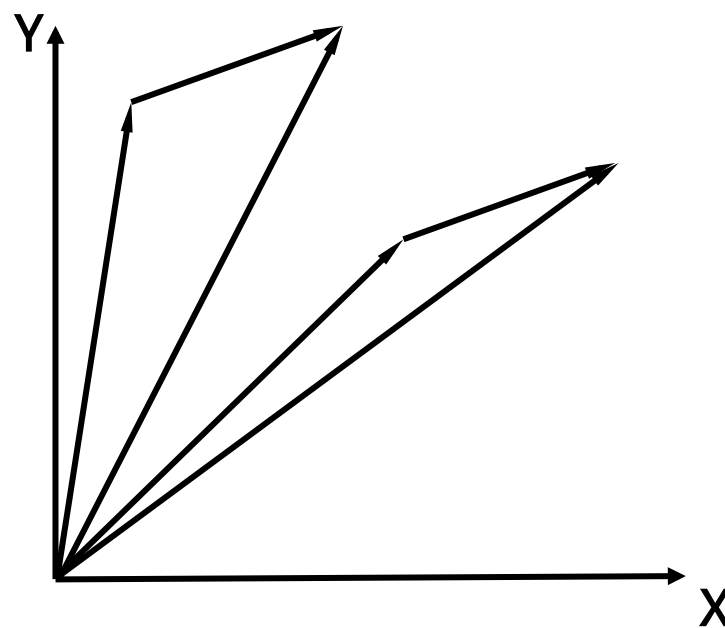




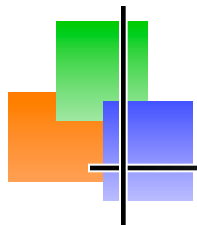
# 基于偏好向量学习的个性化推荐算法



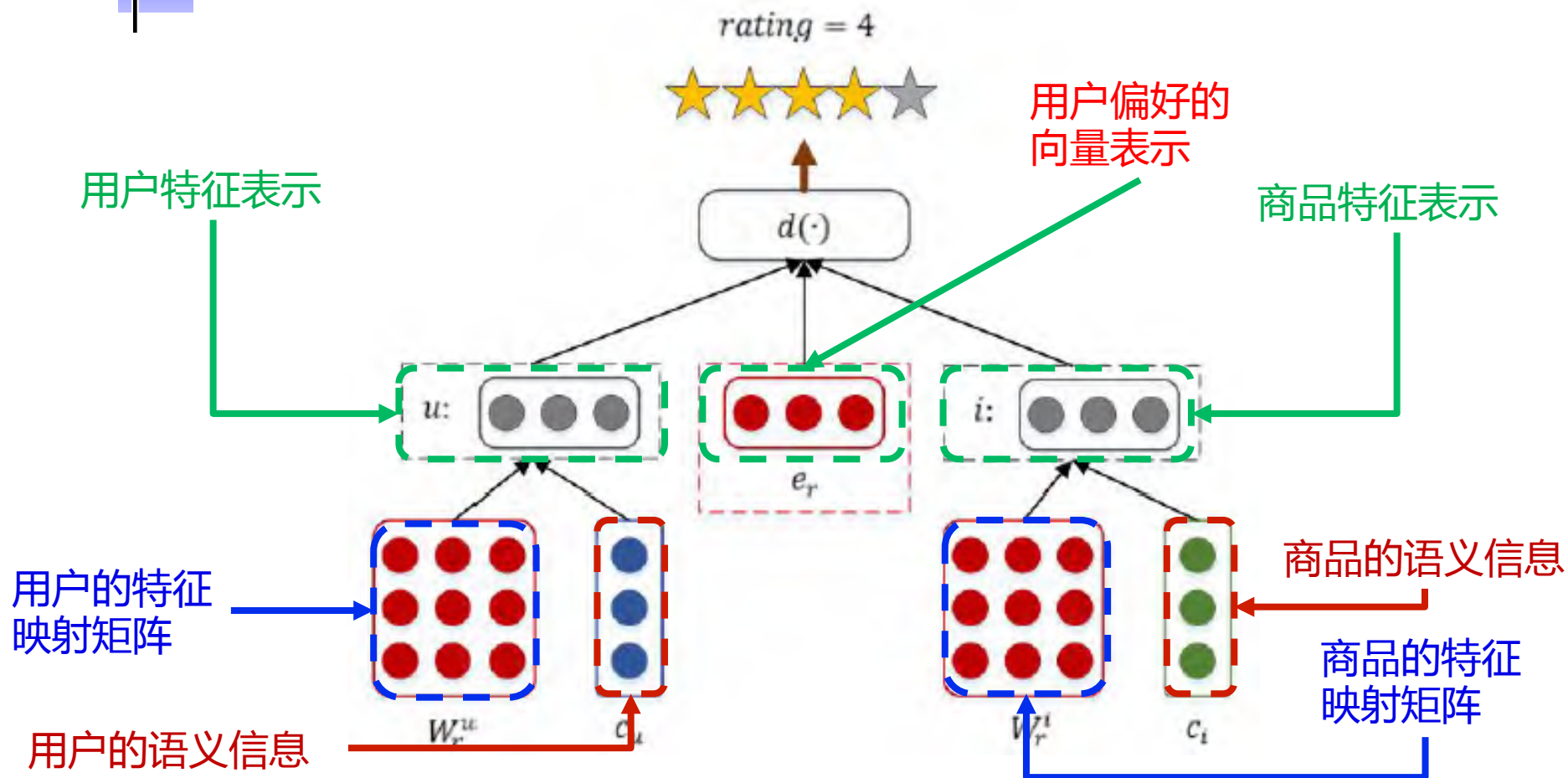
- 假如用户A(为商品X0)打5分，用户B()为商品Y()打5分，那么
- .



$$f_r(h, t) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{\ell_1/\ell_2}$$



# 基于偏好向量学习的个性化推荐算法



$$d_r(u, i) = \|\mathbf{W}_r^u \mathbf{e}_u + \mathbf{e}_r - \mathbf{W}_r^i \mathbf{e}_i\|_{\ell_1 / \ell_2}$$



# 基于偏好向量学习的个性化推荐算法



## 基于偏好向量学习模型在多领域的实验结果

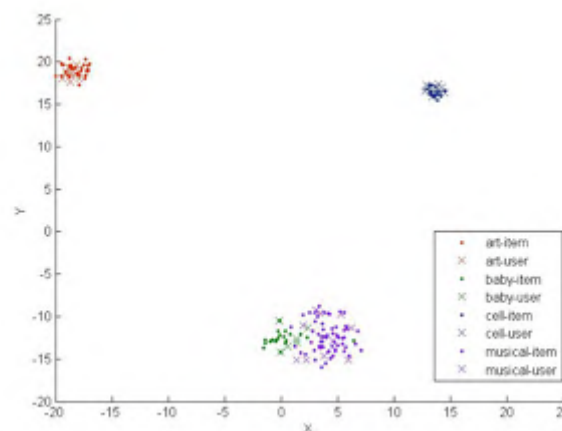
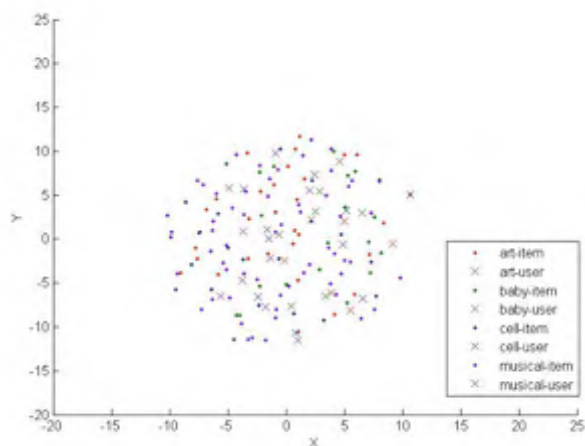
Methods	GlobalAverage	ItemKNN	MF	SVD++	STER		
					random	mf	semantic
Amazon Instant Video	1.073	<b>0.674</b>	0.911	0.777	0.804	0.785	0.688
Arts	1.066	0.899	1.030	0.909	0.865	0.817	<b>0.657</b>
Automotive	1.060	0.874	0.989	0.866	0.816	0.793	<b>0.679</b>
Baby	1.120	0.994	1.097	0.965	1.000	0.991	<b>0.834</b>
Beauty	1.067	0.775	0.906	0.833	0.800	0.758	<b>0.606</b>
Cell Phones & Accessories	1.353	1.247	1.341	1.235	1.448	1.400	<b>1.023</b>
Clothing & Accessories	0.996	0.214	0.351	0.264	0.237	0.219	<b>0.172</b>
Electronics	1.167	0.986	1.119	1.004	1.053	1.033	<b>0.903</b>
Gourmet Foods	1.024	0.901	0.991	0.899	0.778	0.760	<b>0.655</b>
Health	1.092	0.884	1.006	0.918	0.880	0.877	<b>0.733</b>
Home & Kitchen	1.134	0.904	1.040	0.925	0.929	0.930	<b>0.710</b>
Industrial & Scientific	0.855	0.227	0.377	0.243	0.229	0.213	<b>0.181</b>
Jewelry	0.990	0.786	0.946	0.807	0.748	0.815	<b>0.622</b>
Kindle Store	1.018	0.903	0.987	0.909	0.906	0.921	<b>0.744</b>
Musical Instruments	0.987	0.892	0.979	0.885	0.816	0.810	<b>0.718</b>
Office Products	1.147	0.969	1.102	0.987	0.778	0.971	<b>0.756</b>
Patio	1.1729	1.011	1.137	1.015	1.014	1.002	<b>0.814</b>
Pet Supplies	1.107	0.928	1.044	0.947	0.915	0.910	<b>0.751</b>
Shoes	0.900	0.147	0.295	0.197	0.148	0.137	<b>0.107</b>
Software	1.520	1.284	1.488	1.243	1.431	1.350	<b>1.033</b>
Sports & Outdoors	0.997	0.713	0.850	0.747	0.741	0.718	<b>0.586</b>
Tools & Home Improvement	1.087	0.916	1.039	0.928	0.910	0.902	<b>0.752</b>
Toys & Games	1.031	0.905	1.024	0.891	0.878	0.880	<b>0.714</b>
Video Games	1.071	0.944	1.070	0.937	0.936	0.970	<b>0.829</b>
Watches	0.970	0.933	0.970	0.936	0.843	0.843	<b>0.704</b>

Qiu L, Gao S, et al. A Novel Embedding Methods for Recommender System [J]. Knowledge-Based Systems, 2017.

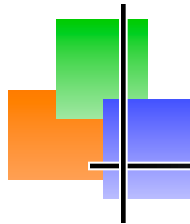
# 基于偏好向量学习的推荐算法



## 矩阵分解模型与偏好向量学习模型的可视化展示



- 不同颜色的点代表了不同类型的商品。左图中所有的点相互紧邻，没有明显的边界，而右图的点明显分为四类，具有明显的边界。
- 由偏好向量学习模型得到的特征更易分类，具有可聚类性。



# 提纲



一. 研究背景

二. 个性化推荐的表示学习模型

A. 基于矩阵分解的Aspect-level推荐算法

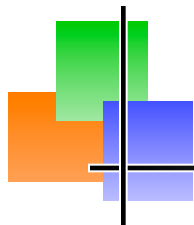
B. 基于偏好向量学习的个性化推荐算法

三. 总结

## □ 基于矩阵分解的Aspect-level推荐算法

- 有机结合了评分信息和评论信息，利用词-文档的表示约束用户和商品向量的表示学习，提高了表示学习的准确性。
- 在精细粒度（Aspect-level）层面上进行推荐，从而提高了推荐的准确度。
- 使原有不可解释模型具有很强的可解释性，用户的评分有其Aspect-level上的依据。

- 基于偏好向量学习的推荐算法
  - 通过将用户偏好（行为）表示为一种分布式向量，可以平滑接入多种行为反馈模式。
  - 通过不同的表示向量初始化方法，可以利用多种先验信息，提高表示学习的效率，提高推荐准确率。
  - 算法执行效率高，运行速度快，便于应用到大规模用户数据。



# 谢谢



高升

北京邮电大学模式识别实验室

北邮本部教三楼803

邮件：[gaosheng@bupt.edu.cn](mailto:gaosheng@bupt.edu.cn)