

5.5 基于内容的推荐

CSDN学院
2017年11月

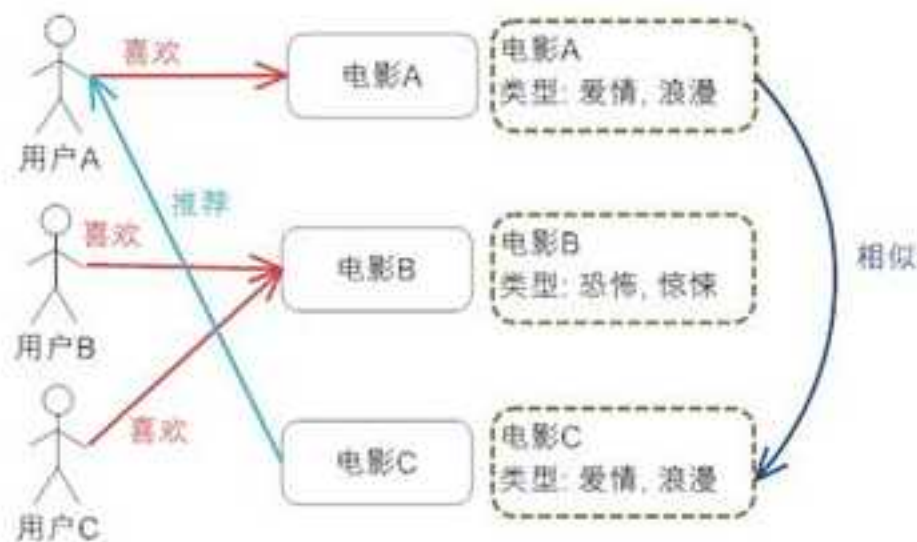
- 推荐系统出现的背景
- 基于协同过滤的推荐
- **基于内容的推荐**
- 推荐系统的评价
- 案例分析

► 基于内容的推荐

- 基于用户喜欢的item的属性/内容进行推荐，无需考虑其他用户的行为
- 根据用户已经选择的对象，计算用户兴趣爱好
- 计算物品的特征，通常使用在文本相关产品上进行推荐
 - Item通过内容(比如关键词)关联:
 - 电影题材:爱情/探险/动作/喜剧/悬疑
 - 标志特征:黄晓明/王宝强...
 - 年代:1995, 2016...
- 计算推荐对象的特征与用户兴趣偏好的匹配程度
- 根据匹配程度排序

► 基于内容的推荐

- 基于用户喜欢的item的属性/ 内容进行推荐，无需考虑其他用户的行为
- 根据用户已经选择的对象，计算用户兴趣爱好
- 计算物品的特征
- 计算物品特征与用户兴趣偏差
- 好的匹配程度
- 根据匹配程度排序



电影A和C都是爱情、浪漫型的电影
用户A喜欢电影A(爱情、浪漫型)
所以给用户A推荐电影C

► 物品画像 (Item Profile)

- 将item表示成一个特征向量
 - 电影：标题、导演、演员、类型...
 - 图像/视频:原数据、标签
- 对文本描述而言，TF-IDF是一种常用的特征
 - 文档集合: $D = \{d_1, d_2, \dots, d_N\}$
 - 词语集合: $T = \{t_1, t_2, \dots, t_n\}$
 - 用向量来表示一篇文档 j : $d_j = (w_{1j}, w_{2j}, \dots, w_{nj})$
 - 其中 w_{kj} 表示第 k 个词 t_k 在文档 j 中的权重，值越大表示越重要，可采用归一化的我TF-IDF

$$\text{TF-IDF}(t_k, d_j) = \underbrace{\text{TF}(t_k, d_j)}_{\text{TF}} \cdot \underbrace{\log \frac{N}{n_k}}_{\text{IDF}}$$

TF：第 k 个词在文档 j 中出现的次数， n_k 包括第 k 个词的文档数量

$$w_{k,j} = \frac{\text{TF-IDF}(t_k, d_j)}{\sqrt{\sum_{s=1}^{|T|} \text{TF-IDF}(t_s, d_j)^2}}$$

► 用户画像 (User Profile)

- 通过用户历史上有过交易(评分、观看...)的item的profiles构建一个模型，然后用训练好的模型预测用户是否会喜欢一个新的item (监督学习任务)
- 理论上机器学习里的分类算法都用于此处
 - k -NN(k 近邻):推荐与用户喜欢的最相似的Item
 - 决策树:当item的属性较少决策树可以产生容易让人理解的结果
 - SVM/Logistic回归
 - 朴素贝叶斯算法
 - . . .

► 基于内容的推荐

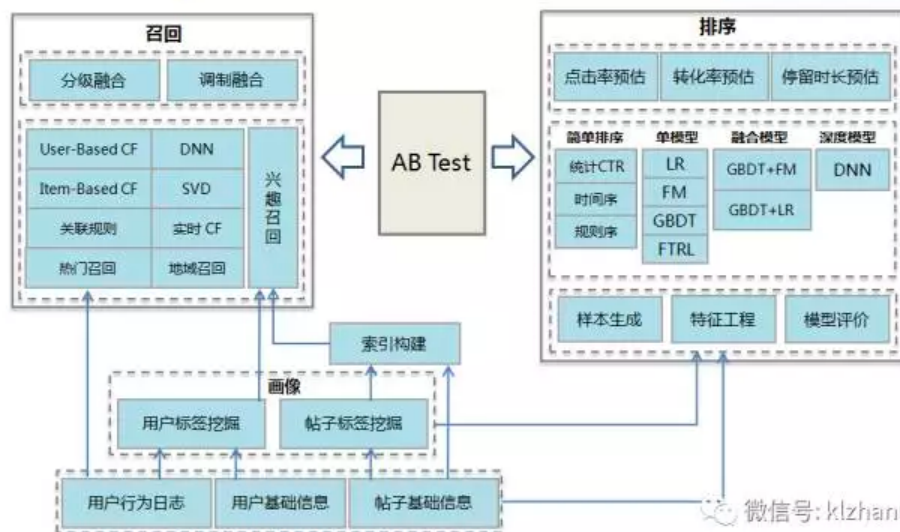
- 优点:
 - 用户之间独立，无需利用很多其他人的数据，抗作弊
 - 可解释性好：产品属性与用户品味匹配
 - 新的item可以立刻得到推荐
- 缺点:
 - item的特征抽取一般很难：特征描述不全可能无法区分不通的item
 - 无法挖掘出用户的潜在兴趣：只与历史兴趣匹配
 - 无法为新用户产生推荐：新用户没有喜好历史，无法建模用户profile

混合推荐

- 各种推荐算法都有不同的优缺点，在实际问题中可以针对具体问题采用各种推荐算法组合进行推荐

推荐算法

让生活更简单 58



THANK YOU



AI100