

协同过滤

推荐+排序

- 1. 如何确定一个用户是否和你有相似品味？
- 2. 如何将邻居们的喜好组成一个排序的目录？

步骤：

- 1. 用户偏好：物品->评分
- 2. 找到相似的用户或物品：比如说第一天，A买了水壶，B也买了水壶；第二天，A买了酸奶，B也买了酸奶；第三天，A买了裤子，是不是也会将裤子推荐给B用户；
- 3. 计算推荐

常见特征：

用户和物品联系到一起：

用户行为	类型	特征	作用
评分	显式	整数量化的偏好，可能的取值是 [0, n]；n 一般取值为 5 或者是 10	通过用户对物品的评分，可以精确的得到用户的偏好
投票	显式	布尔量化的偏好，取值是 0 或 1	通过用户对物品的投票，可以较精确的得到用户的偏好
转发	显式	布尔量化的偏好，取值是 0 或 1	通过用户对物品的投票，可以精确的得到用户的偏好。 如果是站内，同时可以推理得到被转发人的偏好（不精确）
保存书签	显示	布尔量化的偏好，取值是 0 或 1	通过用户对物品的投票，可以精确的得到用户的偏好。
标记标签 (Tag)	显示	一些单词，需要对单词进行分析，得到偏好	通过分析用户的标签，可以得到用户对项目的理解，同时可以分析出用户的情感：喜欢还是讨厌
评论	显示	一段文字，需要进行文本分析，得到偏好	通过分析用户的评论，可以得到用户的情感：喜欢还是讨厌

对于文字，需要对文本进行分析；

基于用户的协同过滤

用户/物品	物品A	物品B	物品C	物品D
用户A	✓	-	✓	✓
用户B	-	✓	✓	-
用户C	✓	✓	-	-

根据用户相似性，来给用户推荐物品

常见问题：

1. 已知用户评分矩阵Matrix R（一般非常稀疏）
2. 推断矩阵中empty cells空格值：基于相似度，找用户，再推荐
3. 新用户，无法找到邻居用户
4. 对于一个物品，所有最近的邻居都在其上没有多少打分

对于问题3和4：

1. 相似度用皮尔逊相关系数
2. 考虑共同打分物品的数目
3. 对打分进行归一化处理
4. 设置一个相似度阈值

用得少，原因：

1. 稀疏问题
2. 数百万的用户计算
3. 人是善变的

应用场景：

1. 实时新闻
2. 突发情况

基于物品的协同过滤

用户/物品	物品A	物品B	物品C
用户A	✓	-	✓
用户B	-	✓	✓
用户C	✓	✓	-

此时计算的是物品和物品之间的相似度，优势：

1. 计算性能高，通常用户数量远大于物品数量，物品可用信息比较多，用户注册时，完善用户信息还是挺麻烦的，但是商品上架时，就会输入各种信息，比如类别，型号等等丰富信息；
2. 可预先保留，物品不善变

应用场景：

1. 图书
2. 电子商务
3. 大众点评 美团
4. 电影

用户冷启动（新用户）：

1. 引导用户把自己的一些属性表达出来
2. 利用现有的开放数据平台
3. 根据用户注册属性
4. 推荐排行榜单（比较稳 大众心理）

物品冷启动（新物品）：

1. 文本分析
2. 主题模型
3. 打标签
4. 推荐排行榜单

评判标准：

1. 均方误差：RMSE（预测值和真正值）；
2. 召回率：推荐列表中，有哪些推荐成功，转化率；
3. 覆盖率：比如说A用户看了一件上衣，那么推荐100件上衣自然要比推荐100件这种款式这种风格的衣服、鞋子等要好； 或者用熵值
4. 多样性：
5. 推荐所带来的广告收益

参考资料

1. <https://www.youtube.com/watch?v=cZp-E98uVFA&list=PLGmd9-PCMLhb9mZoKtAPgKonRopCKR8f9&index=9>