

5.2 基于用户的协同过滤推荐

CSDN学院 2017年11月



▶大纲



- 推荐系统出现的背景
- 基于协同过滤的推荐
 - 基于用户的协同过滤
 - 基于物品的协同过滤
 - 基于模型的协同过滤
- 基于内容的推荐
- 推荐系统的评价
- 案例分析



▶基于协同过滤的推荐



- 基于协同过滤的推荐是最成功的推荐算法,与20世纪90年 代开始研究,并促进了整个推荐系统研究的繁荣
 - 基于"群体智慧"进行推荐
- 协同过滤算法一般分为两类
 - 基于用户的协同过滤
 - 基于物品的协同过滤



▶基于用户的协同过滤



- 基本思想:用户选择某个物品是基于朋友/其他用户的推荐, 给用户推荐与他有相似兴趣的其他用户喜欢的物品
 - 如果一些用户对某些物品的评分相似,则说明这些用户的兴趣偏好相似,那么他们对其他物品的评分也相似
 - 首先找到和目标用户兴趣偏好相似的邻居,然后根据邻居对推荐 对象的评分来预测目标用户对其为评分的推荐对象评分,选择若 干评分高的物品推荐给目标用户



▶基于用户的最近邻协同过滤



- 例: 给用户Alice推荐物品
 - 给定Alice未见过的一个item I, 预测Alice对item I的评分
 - 找到和 Alice过去评分过类似物品且评价过item I的用户
 - 用这些用户对item I 的评分(如平均)作为Alice对item I的评分的预测
 - 对所有Alice未见过的Item评分,取评分最高的物品推荐给Alice

	ltem1	ltem2	ltem3	ltem4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1



▶基于用户的最近邻协同过滤



• 问题

- 怎样度量用户之间的相似性?
- 应该找多少个邻居?
 - 可通过设置相似度阈值或固定邻居数
- 怎样根据邻居的评分生成推荐?

	ltem1	ltem2	ltem3	ltem4	Item5
Alice	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1



▶用户之间的相似性度量



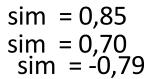
- 用户的特征向量:用户对所有物品的偏好
- 相似性度量:CF中常用的相似性度量为Pearson相关系数

a,b:用户

 $r_{a,p}$: 用户a 对物品p的打分 $sim(a,b) = \frac{\sum_{p \in p} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sum_{p \in p} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in p} (r_{b,p} - \bar{r}_b)^2}$

 $\overline{r_a}$, $\overline{r_b}$: 用户a和b的平均打分

	ltem1	ltem2	Item3	Item4	Item5	
Alice	5	3	4	4	?	
User1	3	1	2	3	3	
User2	4	3	4	3	5	
User3	3	3	1	5	4	
User4	1	5	5	2	1	





▶ 预测



常用预测函数:

$$pred(a, p) = \overline{r_a} + \frac{\sum_{b \in N} \overline{sim(a, b)} * \overline{(r_{b, p} - \overline{r_b})}}{\sum_{b \in N} \overline{sim(a, b)}}$$

- 其他用户对物品i的评分是否高于其平均评分(喜/恶)
- 用户相似性加权
- 加上用户a自身的平均评分偏置



预测



• 例:对用户Alice, $\bar{r}_{Alice} = 4$

• 取最相似的2个相似用户:

- User1 : sim(Alice, User1)=0.85 , $\bar{r}_{User1}=2.25$

- User2 : sim(Alice, User2)=0.7, $\bar{r}_{User1}=3.5$

	ltem1	ltem2	Item3	Item4	ltem5
Alice	5	3	4	4	5
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1





THANK YOU



