

5.3 基于物品的协同过滤推荐

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

► 基于物品的协同过滤

- 基本思想：给用户推荐那些和他们之前喜欢的物品相似的物品
 - 首先找到物品的邻居
 - 再根据用户对邻居的评分，预测用户对目标物品的评分
 - 最后选择若干评分高的物品推荐给用户

► 基于物品的协同过滤

- 物品的特征向量: 所有用户对物品的偏好
- 基于特征向量物品之间的相似度
- 例: 预测Alice对Item5的评分
 - 寻找与Item5相似的物品
 - 根据Alice 对这些物品的评分预测Item5的评分

	Item1	Item2	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

► 物品相似度度量

- 物品相似度可用两个向量夹角余弦表示

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$$

- 对余弦相似度进行调整
 - 考虑每个用户的平均打分 \bar{r}_u
 - U :对物品 a 和 b 打过分的用户的集合

$$\text{sim}(a, b) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

► 物品相似度度量

$$\text{sim}(a, b) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

- $U : \{\text{User1}, \dots, \text{User4}\}$
- $\text{sim}(\text{Item5}, \text{Item1})=0.80, \text{sim}(\text{Item5}, \text{Item2})=-0.91$
- $\text{sim}(\text{Item5}, \text{Item3})=-0.76, \text{sim}(\text{Item5}, \text{Item4})=0.43$

	Item1	Item2	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

- 常用预测函数: 物品相似度加权 用户 u 对物品 i' 的打分, i' 为物品 i 的邻居 (相似物品)

- $$pred(u, i) = \frac{\sum_{i' \in N} sim(i', i) * r_{u, i'}}{\sum_{i' \in N} sim(i', i)}$$

- 用户 u 对物品 i 的邻居的评分是预测基准
- 物品相似性加权

► 预测

$$pred(a, p) = \frac{\sum_{p' \in N} sim(p', p) * r_{a, p'}}{\sum_{p' \in N} sim(p', p)}$$

- 例：用户Alice对物品Item5的打分预测
- 取最相似的2个相似物品：
- Item1： $sim(Item5, Item1)=0.80$, $r_{Alice, Item1} = 5$
- Item4: $sim(Item5, Item4)=0.43$, $r_{Alice, Item4} = 4$

	Item1	Item2	Item3	Item4	Item5
Alice	5	3	4	4	4.65
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

- Item-based算法的预测结果比User-based算法略高。
- Item-based算法可以预先计算好物品的相似度，所以在线的预测速度快。
- 在预测时，所需物品邻居的集合通常很小（因为只需考虑目标用户已经打过分的物品）

THANK YOU



AI100