

# 通俗易懂的协同过滤

原创 尔东木木夕 小羊的梦 1月22日

收录于话题  
#推荐系统 1 #协同过滤 1 #机器学习 1

说起个性化推荐，协同过滤必然是入门级算法了，那么究竟什么是协同过滤，以下将用通俗易懂的demo来讲解，希望能对刚入门推荐的童鞋有所帮助。

- 一 算法核心思想：**物以类聚，人以群分**
1. 跟你喜欢的物品**相似的物品**，你也很有可能喜欢 --基于**物品**的协同过滤推荐（Item-based CF）

2. 跟你喜好**相似的人**喜欢的东西，你也很有可能喜欢 --基于**用户**的协同过滤推荐（User-based CF）
- 二 如何使用协同过滤进行推荐
1. 通过计算物品（或用户）两两之间相似度，即可找出TOP-N相似的人或物品

2. 根据相似的人或物品产生推荐结果

三 举例说明--什么是协同过滤

以下是一个简单的示例，数据集表示一个用户对物品的点击情况：1表示点击，0表示没有点击。

• **User-based CF**

用户/物品	i1	i2	i3	i4	i5
user1	0	1	1	0	0
user2	1	1	1	0	0
user3	1	0	0	1	1
user4	1	1	0	1	1

## 1 计算用户与用户的相似度矩阵

比如有两个用户X和Y，X同学爱好[足球、篮球、乒乓球]，Y同学爱好[网球、足球、篮球、羽毛球]，可见他们的共同爱好有2个，那么他们的相似度可以用： $2/3 * 2/4 = 1/3 \approx 0.33$  来表示。

相似度计算过程如下：

用户	user1	user2	user3	user4
user1	1	$2/2 * 2/3 = 0.67$	0	$1/2 * 1/4 = 0.13$

user2	$2/3 * 2/2 = 0.67$	1	$1/3 * 1/3 = 0.11$	$2/3 * 2/4 = 0.33$
user3	0	0.11	1	$3/3 * 3/4 = 0.75$
user4	0.13	0.33	0.75	1

## 2 根据TOP-N相似的人来进行推荐

此时可能会根据用户原始点击情况，作一些去重或过滤操作。

用户	top2相似用户	初始推荐结果	过滤样本	最终推荐结果
user1	user2,user4	i1i2i3 + i1i2i4i5	i2 , i3	i1, i4, i5
user2	user1,user4	i2i3 + i1i2i4i5	i1, i2, i3	i4, i5
user3	user4,user2	i1i2i4i5 + i1i2i3	i1, i4, i5	i2, i3
user4	user3,user2	i1i4i5	i1, i2, i4, i5	无

说明一下，上面user4没有推荐结果，是因为举例样本数量少，所以刚好出现这种情况，能明白算法原理即可。

### • Item-based CF

#### 1 计算物品与物品的相似度矩阵

物品	i1	i2	i3	i4	i5
i1	1	$2/3 * 2/3 = 0.44$	$1/3 * 1/2 = 0.16$	$2/3 * 2/2 = 0.67$	$2/3 * 2/2 = 0.67$
i2	0.44	1	$2/3 * 2/2 = 0.67$	$1/3 * 1/2 = 0.16$	$1/3 * 1/2 = 0.16$
i3	0.16	0.67	1	0	0
i4	0.67	0.16	0	1	1
i5	0.67	0.16	0	1	1

#### 2 得到TOP-N相似的物品

物品	top2相似物品
i1	i4 , i5
i2	i3, i1
i3	i2, i1
i4	i5, i1
i5	i4,i1

## 3 根据TOP-N相似的物品来进行推荐

用户	已点击物品	初始推荐结果	过滤样本	最终推荐结果
----	-------	--------	------	--------

user1	i2, i3	i3i1 + i2i1	i2,i3	i1
user2	i1, i2 ,i3	i5i1+i3i1+i4i1	i1, i2 ,i3	i4 , i5
user3	i1, i4, i5	i5i1+i5i1+i4i1	i1, i4, i5	无
user4	i1, i2, i4, i5	i5i1+i3i1+i5i1+i4i1	i1, i2, i4, i5	i3

以上就是简单的协同过滤小例子，您觉得希望通俗易懂嘛。嘿嘿。

喜欢此内容的人还喜欢

每天一题leetcode-两数之和（python）

小羊的梦

牛年第一天，省委书记率省委常委集体出境

政知见

【法治热点早知道】2020年全国人均收入最高10个城市，这个省有3个入围！

中国普法