

# 推荐系统

2. 淘宝上有几亿的买家和千万级的商家，我们需要在每个月初给头部的商家（已知商家的名单）推荐一定数量的用户，便于商家有针对性的进行营销活动。

要求：

- 请给出具体的方案和实现方法，包括所用的模型、数据、训练和预测的方法。
- 已有的数据是过去半年所有买家的行为数据，即过去半年的每一天里，单个买家在每个商品上的浏览、收藏、推荐，和购买数量。

总体方案：

方案的总体思路是以商品为媒介，评估用户对各个卖家的评分（兴趣程度），每个卖家挑选兴趣程度最高的前N个用户作为其营销活动的对象。

整个过程可以看做两部分：

- 用户到商品
- 商品到卖家

用户----->商品----->卖家

用户到商品：

这一步评估用户对各个商品的兴趣程度。方法采用基于物品的协同过滤方法。该方法主要分为两步：1、计算各个物品的相似度；2、通过相似度推算用户对所有商品的兴趣度。

1、相似度的计算：

采用文献[1]中相似度的定义：

$$\text{sim}(i_1, i_2) = \frac{\sum_{u \in U} (R_{u,i_1} - \bar{R}_u)(R_{u,i_2} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i_1} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,i_2} - \bar{R}_u)^2}}$$

其中 $i_1, i_2$ 表示两个不同的商品；

$u$ 表示用户；

$U$ 表示同时对 $i_1, i_2$ 有过评分的用户；

$R_{u,i_1}$ 表示用户 $u$ 对商品 $i_1$ 的评分；

$\bar{R}_u$ 表示用户 $u$ 的平均评分；

对所有商品进行两两的相似度计算，就可得到相似度矩阵 $I$ ：

	$i_1$	$i_2$	..	$i_N$
$i_1$	$\text{sim}(i_1, i_1)$	$\text{sim}(i_1, i_2)$	..	$\text{sim}(i_1, i_N)$
$i_2$	$\text{sim}(i_2, i_1)$	$\text{sim}(i_2, i_2)$	..	$\text{sim}(i_2, i_N)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i_N$	$\text{sim}(i_N, i_1)$	..	..	$\text{sim}(i_N, i_N)$

2、通过相似度推算用户对所有商品的兴趣度。

通过物品间的相似度，可以计算用户对未接触到的商品的评分。某一用户 $u$ 对未评分商品 $i_n$ 的评分的估计为：

$$R_{u,i_n} = \frac{\sum_{l=1, l \neq n}^N I_{n,l} R_{u,i_l}}{\sum_{l=1, l \neq n}^N |I_{n,l}|}$$

其中 $I_{n,l}$ 表示矩阵 $I$ 的第 $n$ 行第 $l$ 列。若用户 $u$ 在数据中未接触过商品 $i_l$ ，则 $R_{u,i_l} = 0$ 。通过上面的计算，可以得到关于每个用户对每个商品的评分举证 $UI$ ：

	$i_1$	$i_2$	..	$i_N$
$u_1$	$R_{u_1,i_1}$	$R_{u_1,i_2}$	..	$R_{u_1,i_N}$
$u_2$	$R_{u_2,i_1}$	$R_{u_2,i_2}$	..	$R_{u_2,i_N}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u_M$	$R_{u_M,i_1}$	$R_{u_M,i_2}$	..	$R_{u_M,i_N}$

历史行为转化为评分：

上述基于物品的协同滤波算法需要输入各个用户对一些商品的数值评分。但是用户数据是以历史行为的方式记录，因此需要将历史行为转化为相应的评分。

数据存有半年里用户对每一个商品的浏览、推荐、收藏和购买数量的记录，可以先定义下式作为用户 $u$ 历史行为到评分的映射：

$$R_{u,i} = 0.8 * \text{商品}i\text{购买次数} + (0.6 * \text{商品}i\text{收藏天数} + 0.4 * \text{商品}i\text{推荐次数} + 0.2 * \text{商品}i\text{浏览次数}) / \text{半年天数}$$

一个好的映射需要根据具体的数据进行选择。

商品到卖家：

卖家可以视为商品的集合（缺点：仅仅以商品区分卖家，忽略了卖家服务、信誉、商品价格等重要因素），因此用户对某一个卖家的兴趣程度可以定义为用户对这个该卖家所有商品兴趣程度的总和。基于物品协同滤波算法可以得到用户对所有商品的评分，各个卖家累加它所拥有商品的评分就可得到每个用户对它的评分。

商家可用N\*1维向量S进行表示,S的每个元素取值为0或1,当第m个元素为0表示该商家没有商品 $i_m$ ，1表示有商品 $i_m$ 。所有商家对应的向量构成矩阵**S**:

$$\mathbf{S} = [s_1 \quad s_2 \quad \cdots \quad s_K]$$

关于用户对卖家的评分矩阵为：

$$\mathbf{US} = \mathbf{UI} \times \mathbf{S}$$

**US**为如下矩阵：

	$s_1$	$s_2$	..	$s_K$
$u_1$	$R_{u_1,s_1}$	$R_{u_1,s_2}$	..	$R_{u_1,s_K}$
$u_2$	$R_{u_2,s_1}$	$R_{u_2,s_2}$	..	$R_{u_2,s_K}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u_M$	$R_{u_M,s_1}$	$R_{u_M,s_2}$	..	$R_{u_M,s_K}$

对每一列进行降序排序，就能对每个商家进行用户TopN的推荐。