

# 1. 计算题

1.1 在课件中，我们给出了如下评分矩阵：

		users											
movies		1	2	3	4	5	6	7	8	9	10	11	12
	1	1		3		?	5			5		4	
	2			5	4			4			2	1	3
	3	2	4		1	2		3		4	3	5	
	4		2	4		5			4			2	
	5			4	3	4	2					2	5
	6	1		3		3			2			4	

采用基于用户的评分预测方法（同样采用 2-最近邻），预测用户 5 对于电影 1 的评分，并与课件中给出的基于物品的评分结果进行比较。

先计算用户间相似性

$m(users)$	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{r}_m$ (平均值)	$\frac{4}{3}$	3	$\frac{19}{5}$	$\frac{8}{3}$	$\frac{7}{2}$	$\frac{7}{2}$	$\frac{7}{2}$	3	$\frac{9}{2}$	$\frac{5}{2}$	3	4
$sim(5, m)$	-0.456	-0.949	<u>0.214</u>	<u>0.552</u>	1	-0.158	<u>0.474</u>	<u>0.632</u>	<u>0.474</u>	-0.474	-0.710	0.158

W、 $sim(5, 1)$  为例： $\bar{r}_1 = \frac{1+2+1}{3}$ ， $\bar{r}_5 = \frac{2+5+4+3}{4} = \frac{7}{2}$

$$sim(5, 1) = \frac{(2 - \frac{4}{3}) \cdot (2 - \frac{7}{2}) + (1 - \frac{4}{3}) \cdot (3 - \frac{7}{2})}{\sqrt{(1 - \frac{4}{3})^2 + (2 - \frac{4}{3})^2 + (1 - \frac{4}{3})^2} \cdot \sqrt{(2 - \frac{7}{2})^2 + (5 - \frac{7}{2})^2 + (4 - \frac{7}{2})^2 + (3 - \frac{7}{2})^2}}$$

$$\approx -0.456$$

同理算出其余  $sim(5, m)$ （这里利用 python 求前并得到上述表格）

由2-最近邻 取得用户3和9 (∵ 其余相似度更高的对 movie1 无评分)

$$\therefore \text{pred}(s, 1) = \bar{r}_s + \frac{0.214 \times (3 - \frac{19}{5}) + 0.474 \times (5 - \frac{9}{2})}{0.214 + 0.474}$$
$$\approx 3.6$$

在 PPT 中给出了基于物品时预测结果为 2.6 分，∴ 基于用户的方法预测评分更高。

2.2 试证明：在信息级联 (Information Cascade) 的定义下，信息传播最大化问题的目标函数具有“收益递减”特性，即给定两个集合  $S$ 、 $T$  与集合外的节点  $v$ ，其中  $S \subseteq T$ ，满足：

$$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$$

∵  $S \subseteq T$ , ∴  $f(T) \geq f(S)$ ，对于新节点  $v$ ，有  $f(T) \cap f(v) \geq f(S) \cap f(v)$  ①

①或两边同时  $f(T) - f(S)$  得

$$f(T) - f(S) + f(T) \cap f(v) \geq f(T) - f(S) + f(S) \cap f(v)$$

$$\Rightarrow f(T) - f(S) \geq [f(T) - f(T) \cap f(v)] - [f(S) - f(S) \cap f(v)]$$

$$\Rightarrow f(T) - f(S) \geq [f(T) + f(v) - f(T) \cap f(v)] - [f(S) + f(v) - f(S) \cap f(v)]$$

$$\text{又 } \begin{cases} f(T \cup v) = f(T) + f(v) - f(T) \cap f(v) \\ f(S \cup v) = f(S) + f(v) - f(S) \cap f(v) \end{cases}$$

$$\therefore f(T) - f(S) \geq f(T \cup v) - f(S \cup v)$$

即  $f(S \cup v) - f(S) \geq f(T \cup v) - f(T)$ ，得证

## 2 问答题

2.1 社团推荐是社会网络分析中常见的问题。在**仅考虑网络结构**的情况下，已知网络中存在的若干社团，请结合课程内容中所介绍的推荐技术（包括链接推荐技术）与社团挖掘技术，设计一种给指定节点推荐社团的方法，并简述你的理由。

- 先利用链接预测技术（如RWR指标）为指定结点补全其在原网络结构中的关系  
理由：充分挖掘指定节点在网络中的信息与关系，为下一步社团挖掘提供更多依据。
- 再将各节点利用层次聚类（如Girvan-Newman算法）划分到某个社团，这里实际上因为已知社团数，所以划分过程比较直观简单。
- 当然也可根据Q-Modularity评价指标来筛选  
比如尝试多种链接预测技术（基于多跳关系/有监督学习/仅考虑共同好友数量以及共同好友好友数等），来补全网络，然后比较Q-M值来作优选，使得社团内聚性更强。