

Web作业4

PB18000268 曾勇程

计算题

Web HW4

1. 计算题

记 \bar{u}_i 为第 i 个 user 的平均分。

$$\bar{u}_1 = \frac{1+2+1}{3} = \frac{4}{3} = 1.33$$

$$\bar{u}_2 = \frac{4+2}{2} = 3$$

$$\bar{u}_3 = \frac{3+5+4+4+3}{5} = 3.8$$

$$\bar{u}_4 = \frac{4+1+3}{3} = 2.67$$

$$\bar{u}_5 = \frac{2+5+4+3}{4} = 3.5$$

$$\bar{u}_6 = \frac{5+2}{2} = 3.5$$

$$\bar{u}_7 = \frac{4+3}{2} = 3.5$$

$$\bar{u}_8 = \frac{4+2}{2} = 3$$

$$\bar{u}_9 = \frac{5+4}{2} = 4.5$$

$$\bar{u}_{10} = \frac{2+3}{2} = 2.5$$

$$\bar{u}_{11} = \frac{4+1+5+2+2+4}{6} = 3$$

$$\bar{u}_{12} = \frac{3+5}{2} = 4$$

去中心化后:

	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.33		-0.8		?	1.5			0.5		1	
2			1.2	1.33			0.5			-0.5	-2	-1
3	0.67	1		-1.67	-1.5	-0.5	-0.5	-0.5	0.5	2		
4		-1	0.2		1.5		1				-1	
5			0.2	0.33	0.5	-1.5					-1	1
6	-0.33	-0.8			-0.5		-1				1	

计算相似度:

$$\text{sim}(5, 1) = \frac{0.67 \times (-1.5) + (-0.33) \times (-0.5)}{\sqrt{0.33^2 \times 2 + 0.67^2} \times \sqrt{1.5^2 \times 2 + 0.5^2 \times 2}} = -0.46$$

$$\text{Sim}(5, 2) = \frac{-3}{\sqrt{5} \times \sqrt{2}} = -0.95$$

$$\text{Sim}(5, 3) = \frac{0.8}{\sqrt{5} \times \sqrt{0.8^2 \times 2 + 1.2^2 + 0.2^2 \times 2}} = 0.21$$

$$\text{Sim}(5, 4) = \frac{1.67 \times 1.5 + 0.33 \times 0.5}{\sqrt{5} \times \sqrt{1.33^2 + 1.67^2 + 0.33^2}} = 0.55$$

$$\text{Sim}(5, 5) = 1$$

$$\text{Sim}(5, 6) = \frac{0.5 \times (-1.5)}{\sqrt{5} \times \sqrt{1.5^2 \times 2}} = \frac{-0.75}{\sqrt{5} \times \frac{3\sqrt{2}}{2}} = -0.46$$

$$\text{Sim}(5, 7) = \frac{0.75}{\sqrt{5} \times \sqrt{0.5}} = 0.47$$

$$\text{Sim}(5, 8) = \frac{2}{\sqrt{5} \times \sqrt{2}} = 0.63$$

$$\text{Sim}(5, 9) = \frac{0.75}{\sqrt{5} \times \sqrt{0.5}} = 0.47$$

$$\text{Sim}(5, 10) = \frac{-0.75}{\sqrt{5} \times \sqrt{0.5}} = -0.47$$

$$\text{Sim}(5, 11) = \frac{-3 - 2 - 0.5}{\sqrt{5} \times \sqrt{4 + 2^2 \times 2}} = -0.71$$

$$\text{Sim}(5, 12) = \frac{0.5}{\sqrt{5} \times \sqrt{2}} = 0.16$$

跳过无评分的, 因此和 user 5 最相关的为用户 3 (0.21) 和用户 9 (0.07)

$$\begin{aligned}\therefore \text{pred}(5, 1) &= 3.5 + \frac{0.21 \times (-0.8) + 0.07 \times 0.5}{0.21 + 0.07} \\ &= 3.5985\end{aligned}$$

预测用户 5 对电影 1 的评分为 3.5985 分

与基于物品的评分结果比较: 基于物品的评分结果为 2.6 分, 而基于用户的评分结果为 3.5985 分, 比基于物品的评分结果高了将近 1 分。说明两种评分模型的评分结果对于个体来说有着较大区别。

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。

答: 在已知存在的若干社团的情况下, 基于链接预测推荐技术, 从和该节点有边相连的节点入手 (即这些节点与指定节点有关系), 看看这些节点和哪些社团的关系最大, 这些社团即为该指定节点最可能感兴趣的社团。若边有权重, 权重比较大的边关联的节点的话语权也应该越重, 这是基于若边的权重越重, 该边关联的两个节点的关系越紧密这一点考虑的。

理由：物以类聚，人以群分。好朋友之间兴趣爱好一般也比较相近。在网络结构中，有边相连的两个节点之间一般有着共同的兴趣爱好，故给一个节点推荐社团时，可以从和它相关的节点集合出发，他们最感兴趣的社团也是该指定节点最可能感兴趣的社团。

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：

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

答：节点的影响范围之间可能存在重叠， **v 的影响范围和 T 的影响范围的重叠的范围一定是大于等于 v 的影响范围和 T 的子集 S 的影响范围的重叠的范围的**，即任给一个结点 u ，若它属于 v 的预期激活的节点集合为 $f(\{v\})$ ，同时它属于 S 的预期激活的节点集合为 $f(S)$ ，这它必定属于 S 的父集 T 的预期激活的节点集合为 $f(T)$ 。而 $f(S \cup \{v\})$ 表示的集合为（ S 的预期激活的节点集合）加上（ v 的预期激活的节点集合中和 S 的预期激活的节点集合不重复(叠)的部分）， $f(T \cup \{v\})$ 同理。故 v 的预期激活的节点集合中和 S 的预期激活的节点集合不重复(叠)的部分的数量要大于等于 v 的预期激活的节点集合中和 T 的预期激活的节点集合不重复(叠)的部分的数量，因此有

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