

推荐多样性实验

1. 概述

在之前的 survey 中论述了提升推荐的多样度必然会降低推荐的准确度，需要在其中找到一个平衡点。为了找到这个平衡点，本实验使用了基于 ReRank 方法的几种提升多样性的排序方法，包括 Reverse Rating、Popularity、Average Rating、Absolute Like、Relative Like 等方法，测量了在不同阈值 T_R 下，准确度和多样度的变化；并对比了这几种方法的效果。实验证明，Absolute Like 方法能够在降低少数准确度的情况下，大幅提升多样性。

2. 实验设置

实验数据使用济宁市 2018 年 1 月 1 日到 1 月 27 日女性用户的滑动数据(like/dislike)，由于实验需要预知推荐系统的 Rating 值，而实际的 Rating 值很难获取，因此本实验为每个推荐的 Rating 进行了估算。

这些滑动数据中，每个用户每天滑动的次数各不相同，最少只滑动 1 次，最多高达 7630 次。图 2-1(a)为每人每天滑动数据的 CDF 图，横轴为用户滑动次数，纵轴为少于横轴滑动次数的用户百分比。图 2-1(b)将横轴的取值范围限定在 50 以内。其中滑动次数为 10 次以上的用户占比 75%，20 次以上的用户占比 68%，因此实验取前 20 个作为推荐备选($T_H = 20$)，并假定推荐系统从中取出 10 个作为推荐($N = 10$)，这些值可以随着当前推荐系统的真实情况进行修改。

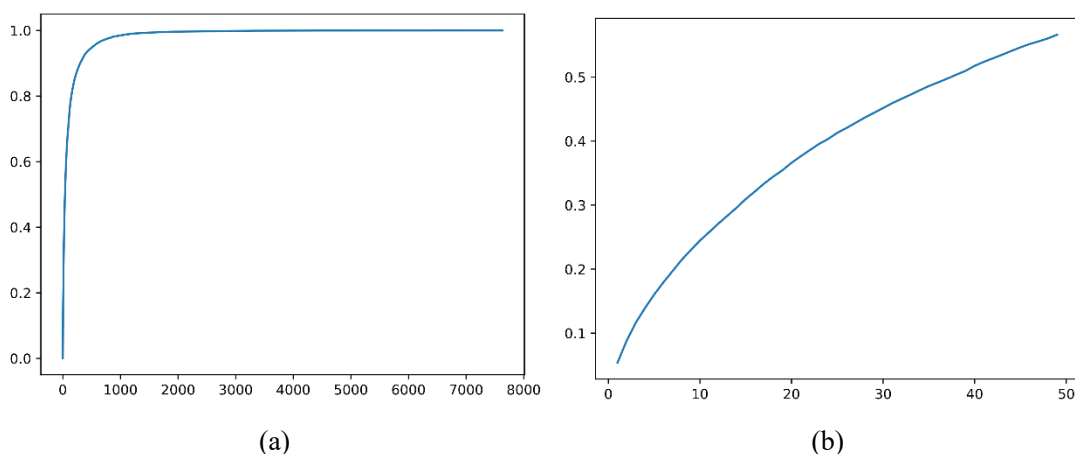


图 2-1 每人每天滑动次数 CDF 图

3. 测量方法

关于准确度的测量，由于缺乏推荐系统完整 Rating 数据（只有滑动数据），在实验中假设原始数据的准确度为 1，实验中测量的准确度均为**相对值**。准确度的测量如下：

$$f(y; \mathbf{u}, \mathbf{I}_u) = \frac{\sum_{i \in y, j \in \mathbf{I}_u \setminus y} [P(i \succ j; u) - P(j \succ i; u)]}{l(n_u - l)}, \quad (6)$$

其中 I_U 表示给用户 U 推荐的所有备选项目， $|I_U| = n_u$ 。并从中抽选 N 个作为推荐， y 为推荐项目集，满足 $y \subseteq I_U$ 。 $P(i \succ j; u)$ 定义如下：

$$P(i \succ j; u) = \begin{cases} 1, & \text{if } r_{u,i} > r_{u,j} \\ 0, & \text{otherwise} \end{cases}$$

函数 $f(y; \mathbf{u}, \mathbf{I}_u)$ 的取值范围为 $[-1, 1]$ 。

多样性的测量使用的是任意两个用户推荐项目的不相似度的平均值，即

$$g(y) = \frac{\sum_{i, j \in y, i \neq j} d(i, j)}{\frac{1}{2}l(l-1)},$$

$d(i, j)$ 为用户 i 的推荐向量和用户 j 的推荐向量的不相似度，定义如下：

$$d(\mathbf{i}, \mathbf{j}) = -\cos(\mathbf{i}, \mathbf{j}) = -\frac{\mathbf{i}^\top \mathbf{j}}{\|\mathbf{i}\| \times \|\mathbf{j}\|}.$$

4. Re-Rank 方法实验

Re-Rank 方法是对于推荐的结果进行重新排序，以提高推荐的多样性。

(a) 按照传统推荐方法，计算 Rating 并排序

	推荐前 10 个给用户																			
Rating	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Popularity	3	9	19	10	8	0	17	5	4	13	2	16	14	7	1	12	6	18	11	15

(b) 按照能提升多样性的方法排序，如这里对 Popularity 从小到大

	推荐前 10 个给用户																			
Rating	14	5	9	19	11	12	3	6	15	18	16	1	4	10	7	0	8	13	2	17
Popularity	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

(c) 调整阈值 T_R 以改善准确度，如 $T_R = 9$ ，即

	推荐前 10 个给用户										$T_R = 9$									
Rating	14	9	19	11	12	15	18	16	10	13	17	5	3	6	1	4	7	0	8	2
Popularity	0	2	3	4	5	8	9	10	13	17	19	1	6	7	11	12	14	15	16	18

图 4-1 Re-Rank 方法改进多样性的过程

如图 4-1 所示为该 Re-Rank 方法改进多样性的过程，分为三步：第一步使用传统的排序方法，计算物品的 rating，并对所有物品的 rating 值进行排序，并、取前 N 个作为推荐项目，这是传统推荐系统的推荐步骤。第二步使用能够提高多样化的排序方法进行排序，该排序方法有以下几种：

Reverse Rating 方法，传统的推荐系统首先计算每个项目 i 对用户 u 的 Rating 值 $R_{u,i}$ ，然后计算每个项目的 rank 值： $Rank_{Standard} = R_{u,i}^{-1}$ ，将每个项目按照 Rank 从小到大排序，取出前 N 个作为推荐。而 Reverse Rating 方法的排序结果与传统方法刚好相反： $Rank_{RevPred} = R_{u,i}$ ，即优先考虑备选中 Rating 值低的项目。

Popularity 方法，该方法首先算出每个项目的被推荐给用户的次数，依照这种原则从小到大排序，即优先考虑备选中不热门的用户，以增加多样性。

Average Rating 方法，计算每个项目被推荐的平均 Rating 值，并从小到大排序。由于数据中没有给出具体的 Rating 值，在实验中假定原先系统推荐给用户第 i 个物品的 Rating 值为 $N - i$ ，作为近似的 Rating 值。

Absolute Likeability 方法，统计每个项目被 like 的数量，并从小到大排序。

Relative Likeability 方法，统计每个项目被 like 的比率，从小到大排序。

第三步，调整 T_R 的值，以改善准确率，限制推荐的项目的 Rating 都满足 $R^*(u, i) \geq T_R$ 。如图中的 Rating=3.65 比 $T_R=3.8$ 小，所以被抛弃，其后一位的项目数顺势补充为第五个推荐项目。

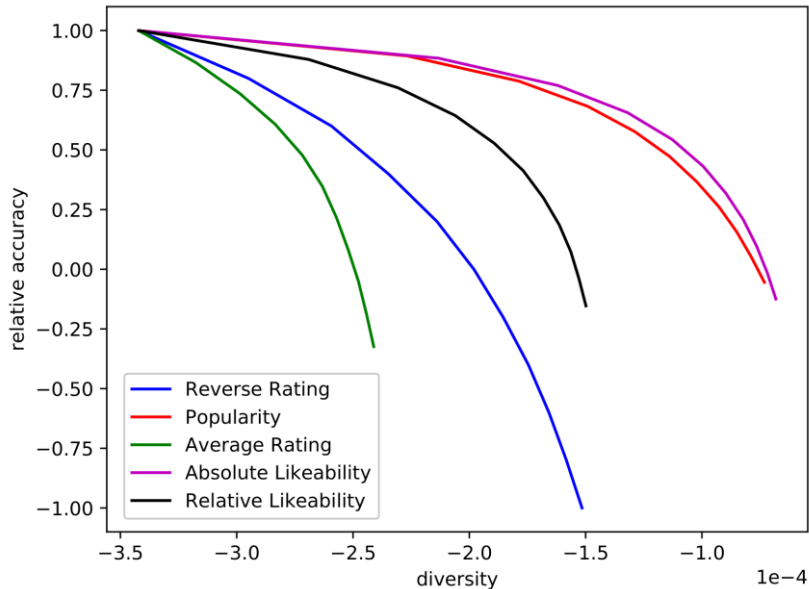


图 4-2 不同排序方法的准确度-多样性曲线图

本实验分别针对上述 5 种排序方法，测量了在不同 T_R 取值情况下，准确度与多样度的变化曲线，如图 4-2 所示：当相对准确度为 1 时为未加入任何优化的系统，其多样性最低，随

随着阈值 T_R 的降低，其准确度出现损失，而多样性增高。表中可以看出，Absolute Likeability 方法能够在损失少量的准确度情况下，实现多样性的大幅度提升，因此该方法推荐在推荐系统中使用以提升多样性。

如图 4-3 所示为使用基于 Absolute Likeability 的 Re-Rank 方法后，对各个项目的推荐次数统计的结果，为了更好地展示结果，图中只取了推荐次数最高的 200 人进行统计。当设定的阈值 T_R 越低，系统的推荐多样性越明显，可以直观地看到长尾效应的改善。

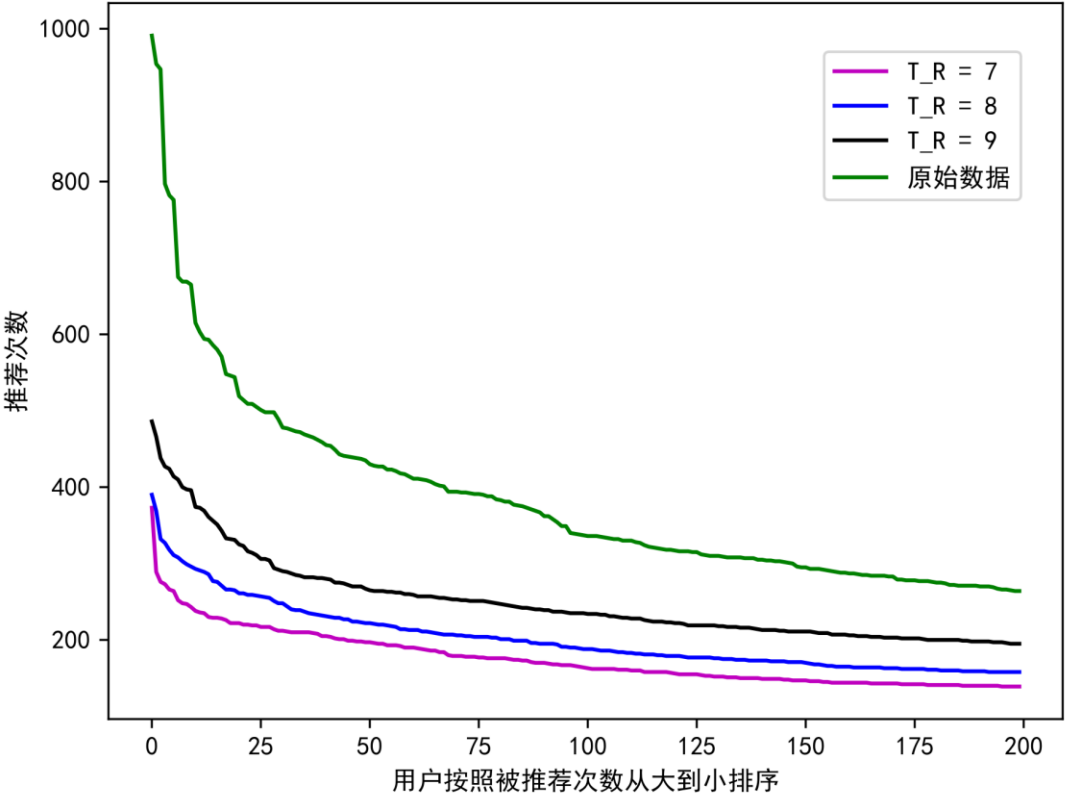


图 4-3 使用基于 Absolute Likeability 的 Re-Rank 方法后，对各个项目的推荐次数统计的结果。