

基于物品的协同过滤实验结果

Author: 李鹏

Date: 2020/05/07

实验任务1输出

- 训练集与测试集大小

```
=====
User-Based-Collaborative-Filtering
=====
```

训练集大小: 980204

测试集大小: 20005

训练集例子:

	user_id	movie_id	ratings
741853	4433	695	3
300817	1782	2770	4
69550	466	30	4

测试集例子:

	user_id	movie_id	ratings
478972	2939	2183	4
698420	4175	2375	3
143115	922	516	1

- 训练集和测试集中的“用户到电影”对应的序列

训练集用户到电影索引举例:

索引值: 4433

对应的列表:

```
[695, 1198, 920, 608, 1358, 3494, 2407, 1676, 2467, 2064, 1090, 25, 1250, 1258, 950, 2248, 2529, 1253, 849, 1201, 1376, 1584, 2797, 1252, 1240, 1617, 2028, 347, 1097, 1805, 3843, 1094, 2100, 2968, 527, 2944, 3365, 3487, 2921, 2951, 1021, 36, 2011, 3836, 750, 1266, 1573, 1210, 908, 3704, 1288, 1320, 3062, 383, 2922, 1294, 198, 1089, 3386, 3476, 3703, 1073, 3730, 293, 541, 1289, 2288, 1215, 2021, 1214, 185, 904, 412, 1086, 6, 1283, 1303, 1219, 2716, 1748, 2640, 3681, 368, 2841, 34]
```

索引值: 1782

对应的列表:

```
[2770, 785, 3176, 2599, 3543, 968, 2034, 2762, 3928, 2901, 1959, 2369, 2987, 250, 3185, 2028, 201, 2683, 2013, 3081, 2794, 1093, 2070, 2541, 858]
```

测试集用户到电影索引举例:

索引值: 2939

对应的列表:

```
[2183, 1394, 2804, 1480]
```

索引值: 4175

对应的列表:

```
2375
```

实验任务2输出

- 物品相似度过程关键数据

得到最终的相似度矩阵:

输出样例:

```
C[u][v]: 124.44995952067062
```

```
movies num or users num: 1697
```

```
movies num or users num: 1702
```

```
sim_matrix: 0.0732274752395855
```

输出样例:

```
C[u][v]: 44.74105799167771
```

```
movies num or users num: 1697
```

```
movies num or users num: 646
```

```
sim_matrix: 0.04273159164337728
```

```
100%|████████████████████| 3705/3705 [00:12<00:00, 291.64it/s]
```

归一化相似度权重:

```
norm_sim_matrix: 0.7299087098306802
```

```
norm_sim_matrix: 0.4259352220376448
```

```
100%|████████████████████| 3705/3705 [00:04<00:00, 895.24it/s]
```

实验任务3输出

- 产生Top-N推荐 (N=10, K=5)

```
需要推荐电影的用户id: 922
推荐的电影:
电影id: 1196 电影名称: Star Wars: Episode V - The Empire Strikes Back (1980)
电影id: 2571 电影名称: Matrix, The (1999)
电影id: 260 电影名称: Star Wars: Episode IV - A New Hope (1977)
电影id: 919 电影名称: Wizard of Oz, The (1939)
电影id: 1265 电影名称: Groundhog Day (1993)
电影id: 1393 电影名称: Jerry Maguire (1996)
电影id: 904 电影名称: Rear Window (1954)
电影id: 380 电影名称: True Lies (1994)
电影id: 3911 电影名称: Best in Show (2000)
电影id: 1952 电影名称: Midnight Cowboy (1969)
```

实验任务4

- 评测模型

N=10, K=5

```
100%|██████████| 118/118 [00:00<00:00, 2405.00it/s]
100%|██████████| 328/328 [00:00<00:00, 4219.82it/s]
Metric: {'Precision': 5.8, 'Recall': 13.61, 'Coverage': 23.71, 'Popularity': 7.227296}
```

N = 10, K=10

```
评测模型:
100%|██████████| 4697/4697 [04:37<00:00, 11.20it/s]
Metric: {'Precision': 6.02, 'Recall': 14.13, 'Coverage': 20.59, 'Popularity': 7.339661}
```

(准确率召回率都提升了, 体现了增加K的优势; 不知道为什么覆盖度降低了; 流行度略有提升也算正常)

N=10, K=20

```
Metric: {'Precision': 6.03, 'Recall': 14.16, 'Coverage': 17.88, 'Popularity': 7.433443}
```

(效果提升不是很明显, 实验耗时明显增加)

N=10, K=50

```
100%|██████████| 4697/4697 [38:09<00:00, 1.67it/s]
Metric: {'Precision': 5.84, 'Recall': 13.7, 'Coverage': 15.12, 'Popularity': 7.515324}
```

(模型效果下降, 应该是K太大, 导致模型正则化太过了)