

基于用户过滤推荐

一、问题描述

| 影片 | 张三 | 李四 | 王五 | 我 |
|----------|----|----|----|-------|
| 《新白娘子传奇》 | 1 | 4 | 3 | 5 |
| 《笑傲江湖》 | 1 | 0 | 4 | 5 |
| 《龙门飞甲》 | 0 | 2 | 1 | 0 |
| 《泰坦尼克号》 | 1 | 1 | 4 | 5 |
| 《流浪地球》 | 0 | 2 | 4 | 值得看么？ |

二、解决方案

1、寻找相似的用户

```
import pandas as pd
from math import sqrt

def get_person_rating(data):
    person_rating = {}
    # 获取每个人对电影的评分

    #如果电影的数据不是按顺序的，使用下面这段代码
    ...
    for j in range(1, 5):
        personj_rating = []
        for i in range(1, 5):
            # 获取第i部电影第j个用户的评分
            personj_rating.append(data[data["movie_id"]==i][data["user_id"]==j]
["rating"].values[0])
        person_rating[j] = personj_rating
    ...

    # 如果电影数据是顺序的，我们就直接取出来，不用在遍历一次movie_id
    for i in range(1, 5):
        person_rating[i] = data[data["user_id"]==i]["rating"].values
    return person_rating

def get_distance(person_rating, index):
    me = person_rating[4] # [5,5,0,5]
    it = person_rating[index]
    sum = 0
    for i in range(0, 4):
```

```
sum = sum + pow(me[i]-it[i],2)
return sqrt(sum)
```

```
# 1、收集数据, 在movie.csv里头
# movie_id,user_id,rating 分别表示电影ID, 用户ID, 评分

# 2、导入数据
data = pd.read_csv("movie.csv")

# 3、获取每一个人的对电影的评分, 要求按照电影ID (1, 2, 3, 4) 顺序
#person_rating = {} # 应该长这样的: {1:[1,2,0,1], 2:[4,0,2,1],3:[3,4,1,4],4:[5,5,0,5]}
person_rating = get_person_rating(data)
#print(person_rating)

# 4、计算我与每一个人的距离
distance = {}
for i in range(1, 4):
    distance[i] = get_distance(person_rating, i)
print(distance)

min_index = 0
min_value = 1000
for key, value in distance.items():
    print("key: %s, value %s" %(key,value))
    if min_value > value:
        min_value = value
        min_index = key
print(min_index)
```

```
{1: 6.928203230275509, 2: 6.782329983125268, 3: 2.6457513110645907}
key: 1, value 6.928203230275509
key: 2, value 6.782329983125268
key: 3, value 2.6457513110645907
3
```

```
pow(3,2)
sqrt(9)
```

2、用相似的用户帮忙推荐