推荐系统

In [1]:

```python
import pandas as pd
from math import pow, sqrt
```

用字典存放所得数据 ¶

In [2]:

```python
file = open('./output/merged.csv', 'r', encoding='utf-8')  # 记得读取文件时加 ‘r’， encoding='UTF-8'
# 读取data.csv中每行中除了名字的数据
data = {}  # 存放每位用户评论的电影和评分
for line in file.readlines():
    # 注意这里不是readline()
    # print(line)
    line = line.strip().split(',')
    # 如果字典中没有某位用户，则使用用户ID来创建这位用户
    if not line[0] in data.keys():
        data[line[0]] = {line[3]: line[1]}
    # 否则直接添加以该用户ID为key字典中
    else:
        data[line[0]][line[3]] = line[1]
```

计算两个用户的相似度

In [3]:

```python
def Euclidean(user1, user2):
    # 取出两位用户评论过的电影和评分
    user1_data = data[user1]
    user2_data = data[user2]
    distance = 0
    # 找到两位用户都评论过的电影，并计算欧式距离
    for key in user1_data.keys():
        if key in user2_data.keys():
            # 注意，distance越大表示两者越相似
            distance += pow(float(user1_data[key]) - float(user2_data[key]), 2)
    return 1 / (1 + sqrt(distance))  # 这里返回值越大，相似度越大
```

找到最相似的k个用户

In [4]:

```python
def top10_similar(userID):
    res = []
    for userid in data.keys():
        if not userid == userID :
            sim = Euclidean(userID, userid)
            res.append((userid, sim))
    res.sort(key=lambda val: val[1], reverse=True)
    return res[:10]
```

In [5]:

```python
tops = top10_similar('1')
print(tops)
```

```
[('userId', 1.0), ('77', 1.0), ('85', 1.0), ('175', 1.0), ('253', 1.0), ('291', 1.
0), ('306', 1.0), ('397', 1.0), ('496', 1.0), ('506', 1.0)]
```

找到最相似的用户看过的电影

In [6]:

```python
def recommend(user, k=5):
    # print(data[user])
    recomm = []
    most_sim_user = top10_similar(user)
    for sim_user in most_sim_user:
        if not sim_user[0] == 'userId':
            # print(sim_user[0])
            items = data[sim_user[0]]
            # print(items)
            for item in items.keys():
                # print(item)
                if item not in data[user].keys():
                    recomm.append((item, items[item]))
    recomm.sort(key=lambda val: val[1], reverse=True)
    return recomm[:k]
```

In [7]:

```python
RECOM = recommend('1')
print(RECOM)
```

```
[('"Incredibles', '5.0'), ('Star Wars: Episode II - Attack of the Clones (2002)',
'5.0'), ('"Lord of the Rings: The Fellowship of the Ring', '5.0'), ('Harry Potter a
nd the Chamber of Secrets (2002)', '5.0'), ('Spider-Man 2 (2004)', '5.0')]
```