

个性化推荐

张岚

December 1, 2016

- I. 取 `index` 为用户 `id`, `column` 为电影 `id`, 从训练集中取对应的用户 `id` 对电影 `id` 的评分, 建立一个新的矩阵, 对于未知的项全定为 0。使用全量数据。

算法 I 数据预处理

输入: `data` 数据

```
I: function PROCData(data)
2:   uids  $\leftarrow$  data.uid.unique
3:   fids  $\leftarrow$  data.fid.unique
4:   df  $\leftarrow$  np.zeros((uids, fids))
5:   for line  $\in$  data do
6:     x  $\leftarrow$  line[1]
7:     y  $\leftarrow$  line[2]
8:     df[x, y]  $\leftarrow$  line[3]
9:   end for
IO:  return df
II: end function
```

2. 使用训练集的数据计算用户的相似度, 再预测测试集中用户对电影的打分, 最后评估准确率。

算法 2 协同过滤

输入: X_{train} 训练数据, X_{test} 测试数据

```
1: function TASK1( $X_{train}, X_{test}$ )  
2:    $sim \leftarrow cosine\_similarity(X_{train})$   
3:    $pred \leftarrow similarity.dot(X_{train})/np.array([np.abs(similarity).sum(axis =$   
    $1)])^T$   
4:    $print \leftarrow RMSE((X_{test}, pred))$   
5: end function
```
