

电影推荐系统技术选型

前端: bootstrap3 框架

后端: django 2.2.1 +djangorestframework (MVC框架)

算法: 协同过滤, 根据用户的打分来进行推荐。从所有打分的用户中找出和当前用户距离最近的n用户, 然后从n个用户打分的电影中找15个当前用户未看过的电影。

最近距离算法通过协同过滤来实现。

推荐算法-协同过滤 - 简书 此项目采用的是皮尔逊相关系数来计算相似度。采取基于用户模型的协同过滤(Neighbor-based Collaborative Filtering)。

数据: python异步爬虫从豆瓣top250抓取数据, 保存到本地csv文件中

数据库: sqlite3

主要功能: 录入图书信息, 用户打分, 电影标签分类, 电影推荐, 电影分享, 电影收藏, 后台管理系统。

整体采用MVC架构, 前端页面通过django template模板来实现, 实现了模板的复用功能。同时前端页面的组织结构较为清晰。

皮尔森距离公式:

$$\rho_{X,Y} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

5.png

基于用户的协同过滤和基于项目的协同过滤区别

协同过滤和基于内容推荐有什么区别? - 知乎

计算物品之间的相似度，然后根据相似度来推荐

物品间的共现矩阵，两个物品同时被n个用户购买

物品间的相似度

$$\omega_{ij} = \frac{|N(i) \cap N(j)|}{\sqrt{|N(i)||N(j)|}}$$

相似度:共现值/N的点赞值*M的点赞值 开根号

推荐值: 相似度*评分