

基于协同过滤的电影精准推荐系统

(毕业设计开题报告草稿)

- 学号：19720152203406
- 姓名：蔡晓聪

研究目标

通过实现基于协同过滤的推荐算法，完成一个电影推荐系统，用户可以在系统上为电影评分，后台再根据用户对过往看过的电影评分来建立一个推荐模型。并依据该推荐模型向用户推荐用户未看过并可能感兴趣的电影。

研究价值

1. 解决信息过载问题

随着移动互联网的流行，信息过载的问题愈加严重。已经有无数的科学家和工程师提出了很多天才的解决方案来解决信息过载的问题。例如我们熟知的分类目录和搜索引擎。著名互联网公司雅虎就以分类目录起家，除此之外中国的Hao123网站等也是，通过将著名的网站分门别类，从而方便用户根据类别查找问题。随着互联网规模的不断扩大，分类目录网站只能覆盖少量的热门网站越来越不满足用户的需求，因此诞生了搜索引擎，以谷歌为代表的搜索引擎可以让用户根据关键词找到自己想要的信息。直到现今仍然是有效解决信息过载的极为有效的解决方案。随着人工智能和机器学习的发展，推荐系统渐渐展露头角，成为解决信息过载的又一利器。特别是当用户无法准确描述自己的需求时更是如此。通过研究推荐系统，能有效地在浩瀚的信息之海中为用户提供更为精准有用的信息，帮助用户解决信息过载问题，因此极有价值。

2. 更好地发掘物品的长尾

美国《连线》杂志主编Chris Anderson在2004年发表了《**The Long Tail**》一文，并于2006年出版了《长尾理论》一书。该书指出，传统的二八原则（80%的销售额来源于20%的热门品牌）在互联网的加入下受到了挑战。互联网条件下，由于传统的“货架”成本变为极端低廉，不热门的商品数量极其庞大，这些长尾商品的总销售额将不容小视。于电影行业而言，对于80%的所谓“冷门”电影，倘若能将他从浩瀚的影库中发掘出来，推荐给极有可能喜欢它的用户，这将挖掘出非常巨大的价值。同时可以很好地规避掉搜索引擎算法带来的马太效应。

3. 电影推荐给推荐系统技术带来了很好的应用场景

精准推荐系统的成功应用有2个条件，第一是存在信息过载，假设用户只有少量的几个选择，那么用户可以很容易地遍历所有项目做出选择。而一百多年来的电影事业的发展，为人类社会带来了无数的影片，因此电影观看选择是存在信息过载的。第二是用户大部分时候没有特别明确的需求，往往用户在闲暇时选择电影观看，常常不是必须要去看特定的电影。很经常就是只是想看电影休闲，但不知道要看什么。此时搜索引擎就显得乏力。因此推荐系统在电影推荐上有十分清晰明确的应用场景。

研究计划

分为以下几个阶段

1. 数据收集和整理

在进行算法设计之前，首先要收集相关的电影数据，利用诸如GroupLens提供的MovieLens这样的开源数据集来作为算法设计时的训练和测试集。除此之外收集电影的数据库为后续开发做准备。

2. 算法的初步设计

根据应用的具体场景，设计出协同过滤算法来训练出推荐模型。并以测试集对模型进行测试，直到达到满意的效果为止。

3. 具体应用的相应开发

完成服务器端的编程，接入训练出的算法模型，提供可供调用的API。并完成前端用户界面的编程，提供图形界面供用户操作，以此积累实际的数据。

4. 完成推荐系统的冷启动

由于一开始时并不具有大量的用户数据，因此需要解决推荐系统的冷启动问题。

5. 书写相关的毕业论文

研究的重难点

1. 协同过滤算法的设计

基于用户行为分析的推荐算法在学术界一般称为协同过滤算法，顾名思义，就是指用户可以齐心协力，通过不断地和网站互动，使推荐列表能够不断过滤掉自己不感兴趣的物品，从而越来越满足自己的需求。协同过滤算法具有多种具体的实现思路，毕业设计的重点要放在设计出更为有效的推荐算法。同时这也是该研究的难点之一。

2. 服务端编程时的根据日志系统完善模型的问题

服务器端记录用户行为的日志之后，该行为日志需要反哺推荐系统的模型。如何根据用户的行为在服务器端实时完善模型。让系统一步一步更懂用户。这也是该研究的一个难点之一。