# 需求调研

**1同类产品对比报告**

1. 豆瓣电影榜单

<https://movie.douban.com/annual/2018?source=movie_navigation#0>

上述为豆瓣2018年度电影榜单。该网页基于豆瓣用户的电影标记数据、排名依据评分、人数和时间综合考虑（反映了千万豆瓣用户的喜好），推荐结果较为准确。

页面依据所推荐的每一部电影划分区域页面，采用分层的方式，每一页介绍了一部电影的概览。同时，页面采用了大篇幅电影海报和宣传片视频作为动态背景，极具感染力。在每一部年度电影的介绍中，都包括了电影的概览、评分详情、电影海报、最高赞评价和相似电影等数据，详细地向观众展现了电影的评价和细节。

以“年度总结”为契机，豆瓣结合平台累积的高质量的评分、影评数据，反映大众最普遍的喜好，极好地呈现了一份推荐榜单。

1. bilibili影视专栏

https://www.bilibili.com/movie/

上述的bilibili电影专栏板块，根据用户平时的观影喜好，针对性地为每一位用户提供个性化的电影推荐。以多样化的小标题分级，为用户提供清晰、简单明了的影视查找功能。除了电影之外，bilibili还将基于用户画像的推荐技术用于其本身弹幕视频的推荐之中。根据用户平时对不同类型视频、不同视频投稿者的观看喜好，进行个性化推荐，从而达到吸引新用户、黏住老用户的目的。

针对于上述2个类似项目的分析，我们的基于用户画像的电影推荐系统将分为2个方面来进行功能需求的概述。

1. 用户画像构建：

用户画像的构建对于推荐模块的运行至关重要。用户画像的信息来源主要由用户的基本属性（静态属性数据）和用户丰富的行为日志（动态画像数据）所组成。画像构建引擎需要结合多部分的数据，动态更新用户画像，从而能兼具个性与时效，不断的更新推荐给用户的内容。

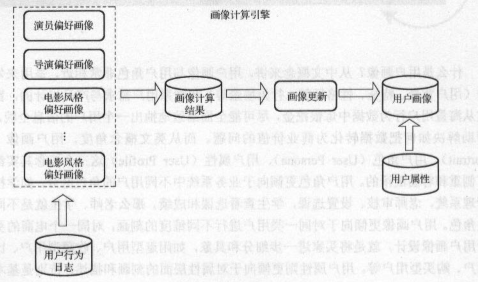


图 1 用户画像的构建

1. 电影推荐模块

电影推荐模块运用相应的推荐算法（基于item内容的、协同过滤等，使用Mahout工具）结合用户画像数据，处理后得到推荐结果。经过融合处理后存储到数据库中供前端调用。

**2 资料收集**

1.行业术语表 （至少5个以上行业术语）

|  |  |
| --- | --- |
| **术语名称** | **简介** |
| 用户画像 | 是用来勾画用户（用户背景、特征、性格标签、行为场景等）和联系用户需求与产品设计的，旨在通过从海量用户行为数据中，尽可能全面细致地抽出一个用户的信息全貌，从而帮助解决如何把数据转化为商业价值的问题。 |
| 协同过滤 | 协同过滤，即利用某兴趣相投、拥有共同经验之群体的喜好来推荐用户感兴趣的信息，个人通过合作的机制给予信息相当程度的回应（如评分）并记录下来以达到过滤的目的进而帮助别人筛选信息，回应不一定局限于特别感兴趣的，特别不感兴趣信息的纪录也相当重要。  协同过滤又可分为评比（rating）或者群体过滤（social filtering）协同过滤以其出色的速度和健壮性，在全球互联网领域炙手可热 |
| Item-Base推荐 | 基于内容的信息推荐方法的理论依据主要来自于信息检索和信息过滤，所谓的基于内容的推荐方法就是根据用户过去的浏览记录来向用户推荐用户没有接触过的推荐项。主要是从两个方法来描述基于内容的推荐方法：启发式的方法和基于模型的方法。启发式的方法就是用户凭借经验来定义相关的计算公式，然后再根据公式的计算结果和实际的结果进行验证，然后再不断修改公式以达到最终目的。而对于模型的方法就是根据以往的数据作为数据集，然后根据这个数据集来学习出一个模型。一般的推荐系统中运用到的启发式的方法就是使用tf-idf的方法来计算，跟还有tf-idf的方法计算出这个文档中出现权重比较高的关键字作为描述用户特征，并使用这些关键字作为描述用户特征的向量；然后再根据被推荐项中的权重高的关键字来作为推荐项的属性特征，然后再将这个两个向量最相近的（与用户特征的向量计算得分最高）的项推荐给用户。在计算用户特征向量和被推荐项的特征向量的相似性时，一般使用的是cosine方法，计算两个向量之间夹角的cosine值。 |
| ALS | ALS算法是基于模型的推荐算法。起基本思想是对稀疏矩阵进行模型分解，评估出缺失项的值，以此来得到一个基本的训练模型。然后依照此模型可以针对新的用户和物品数据进行评估。ALS是采用交替的最小二乘法来算出缺失项的。交替的最小二乘法是在最小二乘法的基础上发展而来的。 |
| Mahout | Mahout 是 Apache Software Foundation（ASF） 旗下的一个开源项目，提供一些可扩展的机器学习领域经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序。Mahout包含许多实现，包括聚类、分类、推荐过滤、频繁子项挖掘。此外，通过使用 Apache Hadoop 库，Mahout 可以有效地扩展到云中。 |
| 用户行为分析 | Mahout 是 Apache Software Foundation（ASF） 旗下的一个开源项目，提供一些可扩展的机器学习领域经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序。Mahout包含许多实现，包括聚类、分类、推荐过滤、频繁子项挖掘。此外，通过使用 Apache Hadoop 库，Mahout 可以有效地扩展到云中。 |

2.行业参考资料（至少3个以上参考网站

https://www.bilibili.com/movie/

<https://movie.douban.com/>

<http://www.iqiyi.com/>

<https://v.qq.com/>

<https://movie.douban.com/annual/2018?source=movie_navigation#0>