

# 推荐系统调研

---

随着网络技术的飞速发展，互联网上存储了海量的数据。在面对如此数量的数据，人们该如何从中选择出想要的信息呢？搜索引擎和推荐引擎都是为此而生，它们的诞生就是为了解决信息过载的问题。

搜索引擎会根据用户的输入，找出相应的数据。而推荐引擎并不知道用户想要什么，用户本身可能也不知道他想要什么，因此推荐引擎只能对用户的行为进行分析，得出想要的结果。

推荐算法是推荐系统的核心，不同的算法各有优劣，我们需要合理的使用。记住，算法虽然是推荐系统的核心，但推荐系统是一个工程问题。一个优秀的推荐系统，数据占 70%，产品占 30%，算法占 10%。

## 推荐算法分类

---

### 基于统计学的推荐算法

就是根据用户的年龄、性别和地理位置等信息推荐他可能感兴趣的物品。

### 基于流行度的推荐算法

常见于各种榜单，根据 PV、UV 等数据找出目前最流行的物品。

### 基于用户（User-based）的推荐算法

当我们要想知道用户对什么物品感兴趣时，我们可以找到与他有相似兴趣的其他用户，然后将他们感兴趣的内容推荐给此用户。这个算法是基于人的，我们只需要找到与用户相似的人群，不需要知道物品更多的附加信息。

它的核心思想就是使用各种方法分析人，找到与他相似的人群。

### 基于内容（Content-based）的推荐算法

和协同过滤算法不同，基于内容的推荐算法更依赖物品的附加信息。它会分析用户感兴趣的物品，然后推荐相似的物品给到用户。

它的核心思想就是使用各种方法分析物品，找到与它相似的物品。

### 基于模型（Model-based）的推荐算法

运用机器学习的思想来进行推荐，通过用户对已有物品的评分来预测他对其他物品的评分，将预测结果中评分较高的物品推荐给用户，是大数据时代最流行的推荐算法。

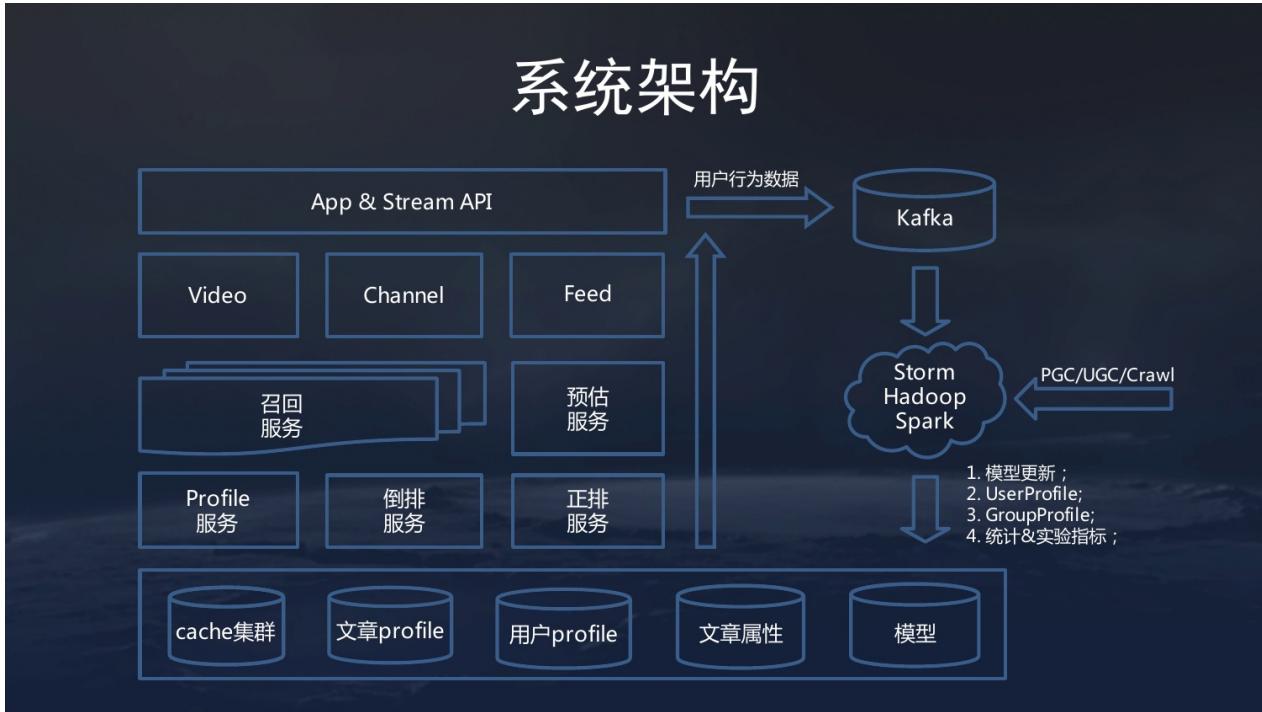
它的核心思想是根据已有的数据集建立预测模型，找到高评分物品。

## 推荐系统

---

今日头条的推荐系统架构：

# 系统架构



## 关键词

冷启动：如何推荐新的物品、如何给新的用户推荐物品、新的推荐系统如何启动。

长尾：马太效应，随着用户行为的累积，推荐系统推荐的物品会集中在用户关注的几个分类中，如何挖掘哪些不那么流行、隐藏的更深的关联物品

召回：根据规则剔除推荐列表中一些过时的物品，相当于一个粗排。

AB 测试：推荐系统的迭代是提高推荐效率的不二选择，离线测试和线上测试都很重要。