

# 1. 项目背景

---

为了能够更好的对尚未上映的电影进行客观评分，供观众进行参考和向观众推荐高质量电影，本项目以豆瓣电影网爬取的2万余部电影作为研究对象，通过计算特征向量的相似度构建电影评分预测模型，进而预测电影评分，为观众客观评分提供参考数据。

## 2. 研究目标

---

应用余弦相似度算法，预测电影评分。

## 3. 研究思路

---

### 3.1 获取源数据

---

爬虫获取豆瓣电影数据2万条

### 3.2 通过余弦相似度算法计算预测评分

---

相似度计算指标（特征） --> 相似度计算 --> 预测评分

## 4. 余弦相似度算法实现

---

### 4.1 筛选相似度计算指标（特征）

---

#### 4.1.1 筛选方法

可视化 + 共现矩阵

#### 4.1.2 筛选结果

类型、导演、编剧、主演、制片国家、语言、上映日期、片长

### 4.2 相似度计算

---

计算各特征相似度后求和

### 4.3 预测评分

---

取与目标电影相似度最高的5部电影，将评分求和后取平均值即为预测评分