第11章 电子商务网站用户行为分析及服务推荐 教案

**课程名称：**Python数据分析与挖掘实战

**课程类别：**必修

**适用专业：**大数据技术类相关专业

**总学时：**112学时（其中理论49学时，实验63学时）

**总学分：**7.0学分

**本章学时**：11学时

# 材料清单

* 1. 《Python数据分析与挖掘实战》教材。
  2. 配套PPT。
  3. 引导性提问。
  4. 探究性问题。
  5. 拓展性问题。

# 教学目标与基本要求

### 教学目标

主要介绍协同过滤算法在电子商务领域中实现用户个性化推荐的应用。通过对用户访问网页日志的数据进行分析与处理，采用基于物品的协同过滤算法进行建模分析，最后通过模型评价与结果分析，得到智能推荐模型。

### 基本要求

1. 熟悉网站智能推荐的步骤与流程。
2. 掌握简单的统计分析方法，运用于网页类型与点击次数的统计。
3. 对某网站数据进行预处理，包括数据去重、数据变换和特征选取。
4. 使用协同过滤算法对某网站进行智能推荐。

# 问题

### 引导性提问

引导性提问需要教师根据教材内容和学生实际水平，提出问题，启发引导学生去解决问题，提问，从而达到理解、掌握知识，发展各种能力和提高思想觉悟的目的。

1. 什么是智能推荐？
2. 生活中常见的智能推荐服务有哪些？
3. 实现智能推荐的算法有哪些？

### 探究性问题

探究性问题需要教师深入钻研教材的基础上精心设计，提问的角度或者在引导性提问的基础上，从重点、难点问题切入，进行插入式提问。或者是对引导式提问中尚未涉及但在课文中又是重要的问题加以设问。

1. 网站的推荐流程是怎么样的？
2. 协同过滤算法除了基于物品的算法外，还有哪些？

### 拓展性问题

拓展性问题需要教师深刻理解教材的意义，学生的学习动态后，根据学生学习层次，提出切实可行的关乎实际的可操作问题。亦可以提供拓展资料供学生研习探讨，完成拓展性问题。

1. 除了协同过滤算法外，能否使用其他算法预测实现网站的智能推荐？

# 主要知识点、重点与难点

### 主要知识点

1. 了解智能推荐服务应用场景。
2. 了解某法律网站现状与数据的基本情况。
3. 熟悉网站智能推荐的步骤与流程。
4. 分析原始数据用户点击的网页类型，得到统计分析结果。
5. 根据原始数据用户浏览网页次数的情况进行统计分析。
6. 通过原始数据用户在浏览页面时的情况得到网页排名的统计分析。
7. 识别翻页的网址，并对其进行还原，然后对用户访问的页面进行去重操作。
8. 对数据进行手动网址分类，对处理后的数据进行特征选取。
9. 基于物品的协同过滤算法，计算出物品之间的相似度。
10. 根据物品的相似度和用户的历史行为给用户生成推荐列表。
11. 对模型进行评价，判断推荐系统的好坏。

### 重点

1. 网站智能推荐的步骤与流程。
2. 统计网页整体流量状况。
3. 基于物品的协同过滤算法，计算出物品之间的相似度。
4. 根据物品的相似度和用户的历史行为给用户生成推荐列表。
5. 对模型进行评价，判断推荐系统的好坏。

### 难点

1. 基于物品的协同过滤算法，计算出物品之间的相似度。
2. 根据物品的相似度和用户的历史行为给用户生成推荐列表。

# 教学过程设计

### 理论教学过程

1. 了解智能推荐服务应用场景。
2. 了解某法律网站的基本情况。
3. 分析原始数据用户点击的网页类型，得到统计分析结果。
4. 根据原始数据用户浏览网页次数的情况进行统计分析。
5. 通过原始数据用户在浏览页面时的情况得到网页排名的统计分析。
6. 清除数据探索分析过程中发现与目标无关的数据。
7. 识别翻页的网址，并对其进行还原，然后对用户访问的页面进行去重操作。
8. 对数据进行手动网址分类，对处理后的数据进行特征选取。
9. 基于物品的协同过滤算法，计算出物品之间的相似度。
10. 根据物品的相似度和用户的历史行为给用户生成推荐列表。
11. 对模型进行评价，判断推荐系统的好坏。

### 实验教学过程

1. 使用pymysql库中的pm连接数据库。
2. 统计101、107和199等网页类型。
3. 统计用户在网页的点击次数。
4. 分析网页排名。
5. 删除不符合规则的网页。
6. 还原翻页网址。
7. 划分正确的网页类型。
8. 将处理后的数据转换成0-1二元型数据。
9. 构建智能推荐模型。
10. 评价智能推荐模型。

# 教材与参考资料

### 教材

张良均，谭立云，刘名军，江建明．《Python数据分析与挖掘实战（第2版）》[M]．北京：机械工业出版社．2019．