# 电影信息智能推荐平台的构建案例教学指导手册

**教学适用的课程：**《大数据架构与设计》

**教学适用的专业：**电子信息

**教学目的与用途：**理解互联网推荐系统的应用；掌握大数据预处理、数据分析相关知识点；掌握常用推荐算法的应用。

**教学内容：**

（1）理论依据：本案例基于大数据技术进行互联网推荐系统的构建。案例整合了常用的数据挖掘与机器学习方法，进行网络电影信息数据的爬取以及预处理，同时采用经典推荐算法实现了电影个性化智能推荐。

（2）涉及知识点：网络数据爬取；数据预处理；推荐算法。

（3）分析路径：首先明确互联网推荐系统设计的实际意义，并以构建电影信息推荐平台作为具体问题。通过Hadoop 和Spark搭建服务平台，此后操作均基于该平台。随后爬取热门电影网站的电影信息数据进行存储，并对其中错误数据及缺失数据进行预处理。最后采用常用推荐算法如SVD算法和ALS算法进行电影个性化智能推荐。

**启发思考题：**

推荐过程中需要计算电影特征向量之间的相似度评分，除了通过余弦相似度进行度量，还能想到其他的度量方法吗？（参考答案：欧几里得距离；Pearson相关系数；Jaccard系数）

**建议课堂计划：**

1. 时间安排：3课时
2. 学生学习准备：了解Hadoop 、Spark、网络爬虫以及常用推荐算法理论知识。
3. 分组及讨论内容：分组讨论启发思考题。

案例开场白：推荐系统根据用户需求､兴趣等，通过推荐算法从海量数据中挖掘出用户感兴趣的项目（如信息､服务､物品等），并将结果以个性化列表的形式推荐给用户。目前，推荐系统在很多领域得到了成功应用，包括电子商务（如Amazon､eBay､Netflix､阿里巴巴等）､信息检索（如iGoogle､MyYahoo､百度等）､社交网络（Face-book､Twitter､腾讯等）､位置服务（如Foursquare､Yelp､大众点评等）､新闻推送（如GoogleNews､GroupLens､今日头条等）等各个领域。

1. 结束总结：这节课我们以构建电影信息推荐平台为例。通过Hadoop 和Spark搭建服务平台。随后爬取热门电影网站的电影信息数据进行存储，并对其中错误数据及缺失数据进行预处理，最后采用常用推荐算法如SVD算法和ALS算法进行电影个性化智能推荐。我们通过平时课程所学知识实现了互联网中的推荐系统，课后感兴趣的同学可以搜集相关资料，进一步了解大数据在其他场景的应用，相信你们会收获更多。
2. 案例引导建议：在教师简要分析应用场景后，鼓励学生自主考虑如何使用所学知识解决该场景问题。

**参考文献：**

[1] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

[2] 黄立威,江碧涛,吕守业,刘艳博,李德毅.基于深度学习的推荐系统研究综述[J]. 计算机学报,2018,41(07):1619-1647.

**其他教学支持材料：**

1. 本案例提供配套的PPT、视频、数据集与代码等，发布于Github，链接为：https://github.com/Wanghui-Huang/CQU\_bigdata。

2. 本案例涉及数据预处理以及多种机器学习算法，建议使用python语言进行编写，推荐的工具包有requests（网络请求库），BeautifulSoup（HTML解析库），scikit-learn（机器学习库）。