# 网易云音乐评论情感分析研究案例教学指导手册

**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：**熟悉Hadoop、Spark等大数据技术栈；掌握分布式大数据处理环境的搭建；掌握大数据处理与分析的基本流程。

**教学内容:**

（1）理论依据：本案例基于大数据技术对网易云音乐评论进行情感分析。案例整合了常用的数据挖掘与深度学习方法，进行数据爬取、数据预处理，实现了网易云音乐评论情感分析，为企业向用户推荐喜好的音乐提供了决策。

（2）涉及知识点：Hadoop&Spark技术、分布式爬虫、深度学习、数据可视化。

（3）分析路径：首先在Hadoop&Spark基础环境下采用Scrapy-redis分布式爬虫框架爬取网易云评论和用户数据，存储在mongodb里面，之后采用snowNLP作为对照，直接对数据情感进行打分，接着搭建GRU和CNN模型来对数据进行情感分析和比较，其中GRU以及CNN是通过手工标注的数据利用tensorflow实现并训练得到，之后使用Spark框架来对数据进行整合和分析，最后将分析结果导出后使用Echarts将数据进行可视化展示，部署到云服务器上。

**启发思考题:**

在文本情感分类领域，什么情况下采用GRU，什么情况下采用CNN？

（参考答案：当句子的情感分类是由整个句子决定的时候，GRU会更容易分类正确， 当句子的情感分类是由几个局部的key-phrases决定的时候，CNN会更容易分类正确）

**建议课堂计划：**

1. 时间安排：3课时
2. 学生学习准备：了解常用的大数据技术栈以常见的数据预处理、情感分析方面的理论知识。
3. 分组及讨论内容：分组讨论启发思考题。
4. 案例开场白：随着络技术的发展，越来越多互联网用户在各种网站上分享经验和发表评论，这些评论文本表达了评论者的观点和情感，蕴含了巨大的商业价值。借助自动化工具对互联网评论文本进行情感分析，能够发现用户情感的演化规律，挖掘其中的潜在信息。本节课将学习如何对网易云音乐评论进行情感分析。
5. 结束总结：这节课我们在网易云音乐评论情感分析任务上，使用了分布式爬虫技术，熟悉了snowNLP模型、GRU模型和CNN模型，学习了如何使用Spark框架来对数据进行整合和分析以及如何利用Echarts对结果进行可视化处理。我们通过平时课程所学知识去解决了实际社会生活中的问题，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行业的应用，相信你们会收获更多。
6. 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]吴宇鹏.分布式网络爬虫技术的研究与实现[J].电脑编程技巧与维护,2020(11):9-10+19.DOI:10.16184/j.cnki.comprg.2020.11.003.

[2]张胜敏,王爱菊.基于Python的分布式多主题网络爬虫的研究与设计[J].开封大学学报,2021,35(01):93-96.

[3]史国举.数据可视化技术在大数据分析领域的应用及发展研究[J].无线互联科技,2021,18(18):96-97.

**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. 本案例涉及到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pandas（数据读取与预处理库），scikit-learn（机器学习算法库），Echarts（可视化绘图库）。