

NLP项目大纲（12课时）

陈晓理

目的：熟悉自然语义情感分析问题的操作流程，理解相关NLP算法，加深对机器学习算法的理解，熟练python数据操作

方法：结合案例实践与理论指导。案例链接：<https://www.lintcode.com/ai/movie-review-recognition/overview>

大纲：

1. 影评项目简介，梳理自然语义情感分析的操作流程，安装相关工具。
2. NLTK与结巴分词的原理与使用，建立语料库
3. Python 数据操作: pandas, numpy,
4. 语言文本等进一步处理与特征提取（1）：TF-IDF的原理
5. 语言文本等进一步处理与特征提取（2）：Word2Vec的概念与应用
6. 机器学习的分类问题（1）：适用范围，评判标准（ROC, AUC等）以及与概率统计的关联
7. 机器学习的分类问题（2）：拟合与过拟合，交叉检验
8. 分类模型原理，应用及效果（1）：朴素贝叶斯
9. 分类模型原理，应用及效果（2）：决策树与随机森林
10. 分类模型原理，应用及效果（3）：SVM
11. 分类模型原理，应用及效果（4）：循环神经网络
12. 组合模型以及前沿发展

注：实际教学时会根据教学情况调整进度与内容

参考书目：

Joel Grus. <Data Science from Scratch. First Principles with Python>. O.Reilly

周志华. 《机器学习》. 清华大学出版社

李航. 《统计学方法》. 清华大学出版社

黄文坚 唐源. 《TensorFlow 实战》. 电子工业出版社