

项目思路

- nlp常见任务流程:

- ◆ 数据准备和定义

- ◆ 数据观察和清洗

观察文本长度，词频分布等特征和标注；去停用等不相关词，词干化（stemming），词形还原（lemmatization）等

- ◆ 特征工程/数据表示

词形：Bag of Words, tf-idf等 词义：word2vec, 基于wordnet的相似度等

- ◆ 模型选择

传统机器学习方法，深度学习方法

- ✓ 验证和错例反思观察

- ◆ 集成方法

数据的集成，分类器的集成

作业点评

• 优秀作业奖:

综合排行榜成绩和项目报告结果，优秀作业奖如下:

★ Newbee: 常永炷

思路广阔，基本覆盖了传统机器学习方法和深度学习方法

★ 要没时间了: 陈文帆

项目思路和逻辑清晰，特征工程做得不错

• 鼓励奖:

✿ 王泽群

作为一名Nlp新手，短期内取得这样的成绩，精神可嘉。

项目文件不要再贴代码，多花时间去思考下工具背后的含义

Ta1相关业务

- Nlp

- ★ 托福作文批改

每年50w左右的作文量;e-rater;个性化修改;

- ★ 教师上课内容分析

- speech:

- ★ 教育场景下的语音识别技术

- ★ 语音合成

- ★ 语音情感分析