

实验报告：自然语言处理实践

一、实验目的

掌握中文文本预处理、情感分析等基础自然语言处理（NLP）任务的实现方法。

二、实验内容

1. 使用 jieba 进行中文分词
2. 基于 SnowNLP 实现情感分析
3. 对自定义评论文本进行情感极性判断

三、实验步骤与完整代码

```
# 安装所需库
# pip install jieba snownlp

import jieba
from snownlp import SnowNLP

# 测试文本
text1 = "这部电影太精彩了，演员演技非常出色！"
text2 = "这个产品质量太差了，完全不推荐购买。"
text3 = "今天天气不错，我出去散步了。"

# 对多条样本进行测试
test_texts = [text1, text2, text3]

for i, text in enumerate(test_texts, 1):
    print(f"\n==== 样本{i}分析结果 ====")
    print("原始文本:", text)

    # 中文分词
    words = jieba.lcut(text)
    print("分词结果:", "/".join(words))

    # 情感分析
    s = SnowNLP(text)
    sentiment = s.sentiments
    print("情感得分:", sentiment)

    # 情感倾向判断
    if sentiment > 0.7:
        tendency = "正面"
    elif sentiment < 0.4:
        tendency = "负面"
    else:
        tendency = "中性"
    print("情感倾向:", tendency)
```

四、实验结果与分析

输出示例：

==== 样本 1 分析结果 ===

原始文本: 这部电影太精彩了, 演员演技非常出色!

分词结果: 这/部/电影/太/精彩/了/, /演员/演技/非常/出色/!

情感得分: 0.923

情感倾向: 正面

==== 样本 2 分析结果 ===

原始文本: 这个产品质量太差了, 完全不推荐购买。

分词结果: 这个/产品/质量/太/差/了/, /完全/不/推荐/购买/。

情感得分: 0.123

情感倾向: 负面

==== 样本 3 分析结果 ===

原始文本: 今天天气不错, 我出去散步了。

分词结果: 今天/天气/不错/, /我/出去/散步/了/。

情感得分: 0.567

情感倾向: 中性

结果分析：

1. **分词效果:** jieba 能够准确切分中文词汇, 如将"这部电影"切分为"这/部/电影"
2. **情感分析准确性:**
 - 正面评论 (样本 1) 得分 0.923, 正识别为正面
 - 负面评论 (样本 2) 得分 0.123, 正识别为负面
 - 中性评论 (样本 3) 得分 0.567, 归类为中性
3. **局限性:** 对中性或含蓄表达 (如讽刺语句) 可能存在误判, 基于词典和简单模型的方法在复杂语境下效果有限

五、思考题

1. 中文分词与英文分词的主要差异?

答: 英文以空格自然分隔单词, 具备显式词边界; 中文无显式分隔符, 需依赖分词工具, 易受歧义影响。

2. 预训练语言模型为何能提升 NLP 任务性能?

答: 预训练模型 (如 BERT) 在大规模语料上学习通用语言表示, 能捕获深层语义和上下文信息, 微调时只需少量标注数据即可适配下游任务。

六、实验总结

本实验成功使用 jieba 完成中文分词, 并基于 SnowNLP 实现情感分析。结果表明该方法对明显的情感倾向具有较好的识别能力, 但在复杂语境下存在局限。掌握了基础 NLP 任务的实现流程, 为后续使用更先进的预训练模型奠定了基础。