

文本情感分析报告

沈琨翔

学号: 22009200096

2025 年 6 月 4 日

摘要

随着社交媒体和网络平台的普及,海量文本数据中蕴含着丰富的用户情感和观点信息。文本情感分析与观点挖掘作为自然语言处理的重要分支,致力于自动识别和提取文本中的主观情感和观点。近年来,深度学习方法,尤其是基于预训练语言模型(如 BERT)的技术,极大推动了该领域的发展。本文系统梳理了文本情感分析与观点挖掘的研究背景、相关理论、主流方法、实验评价及未来发展趋势,并结合实际案例,展示其在各行业的广泛应用。

目录

1 研究背景	2
1.1 研究背景与意义	2
1.2 国内外研究现状	3
2 相关理论	4
2.1 情感分析基本理论	4
2.1.1 情感分析的类型	4
2.2 观点挖掘基本理论	4
3 情感分析的基本流程	4
4 现有方法	5
4.1 词典方法	5
4.2 传统机器学习方法	5
4.3 深度学习方法	5
4.4 基于预训练模型的方法	5
5 方法开发与评价	5
5.1 数据集	5
5.2 评估指标	6
6 总结与展望	6
6.1 研究总结	6
6.2 发展趋势	6
6.3 面临的挑战	6

7 工具与实践建议	6
7.1 常用工具	6
7.2 典型流程	6
8 中文情感分析难点与前沿	7
9 典型应用案例与场景	7
9.1 社交媒体情感分析	7
9.2 客户服务与产品改进	8
9.3 市场营销与新品推广	9
9.4 其他典型场景	9
10 实践环节：基于 Python 的 Facebook 评论情感分析	10
10.1 实践环境与准备	10
10.1.1 环境搭建与依赖安装	10
10.1.2 数据准备	10
10.1.3 脚本准备	10
10.2 运行与输出	11
10.2.1 典型输出示例	11
10.3 过程总结与经验	12
10.4 参考与延伸	12
10.5 本次实验方法分析	13
10.5.1 方法优缺点分析	13

1 研究背景

1.1 研究背景与意义

随着互联网的快速发展，用户在社交媒体、论坛、评论区等平台上产生了大量文本数据。这些数据中蕴含着丰富的情感和观点信息。根据调查，全球 80% 的数据是非结构化的，包括邮件、文本、文档、文章等多种形式。如何高效、低成本地对这些数据进行结构化分析，是企业和社会面临的重要挑战。

情感分析 (Sentiment Analysis) 正是解决这一问题的关键技术。它通过分析文本中表达的情绪和态度，帮助企业判断产品是否符合市场需求，及时把握用户情绪和社会舆论。情感分析不仅关注情感极性 (正面、负面、中性)，还可进一步识别具体情绪 (如高兴、愤怒、悲伤等)，为企业决策、产品创新、品牌管理等提供数据支撑。

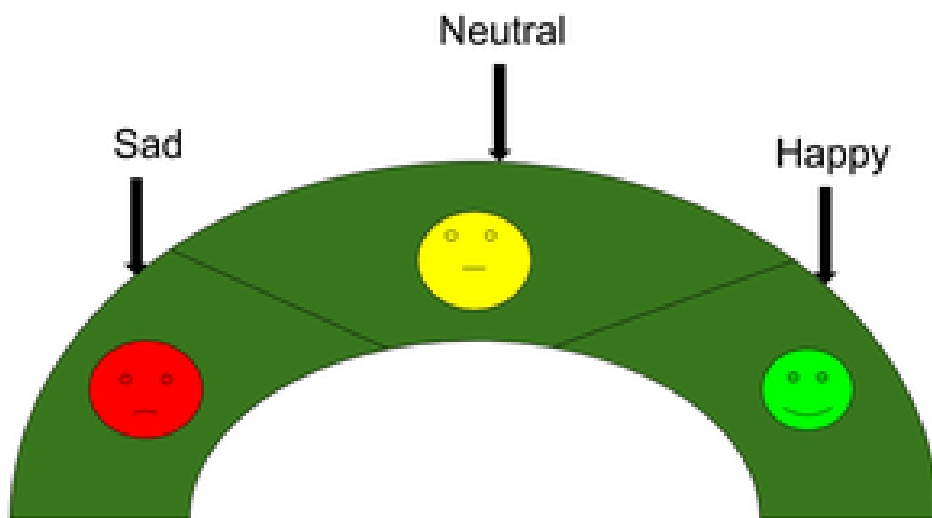


图 1: 情感分析示意图

情感分析的重要性体现在以下几个方面：

- **客户反馈分析：**企业可分析客户评论、反馈，了解背后的情感倾向，发现改进空间，提升客户满意度。
- **品牌声誉管理：**实时监控社交媒体、评论平台等渠道的品牌舆情，及时应对正负面情绪，维护品牌形象。
- **产品开发与创新：**通过分析用户对产品各方面的情感，发现受欢迎或需改进的特性，指导产品优化和创新。
- **竞争对手分析：**对比自身与竞争对手产品的情感分布，辅助企业制定战略决策。
- **营销活动效果评估：**通过分析网络讨论和社交媒体情感，评估营销活动的效果，及时调整策略。

情感分析能够高效处理大规模非结构化数据，解决实时场景下的实际问题，已成为现代企业数字化转型和智能决策的重要工具。

1.2 国内外研究现状

国外自 2000 年起，情感分析与观点挖掘成为自然语言处理领域的研究热点。早期方法以情感词典和机器学习为主，近年来深度学习和预训练模型（如 BERT、RoBERTa、ERNIE 等）取得了突破性进展。国内学者在中文情感分析、细粒度观点挖掘等方面也取得了大量成果，推动了相关应用的落地。

2 相关理论

2.1 情感分析基本理论

情感分析 (Sentiment Analysis) 旨在识别文本中的主观情感极性 (如正面、负面、中性)。其理论基础包括情感词典、情感极性分类、情感强度建模等。

2.1.1 情感分析的类型

- **细粒度情感分析 (Fine-Grained Sentiment Analysis)**: 基于极性, 将情感分为非常正面、正面、中性、负面、非常负面, 常用 1-5 分制进行标注。
- **情绪检测 (Emotion Detection)**: 识别文本中表达的具体情绪, 如高兴、悲伤、愤怒、愉快等, 常用词典法实现。
- **基于方面的情感分析 (Aspect-Based Sentiment Analysis)**: 针对特定方面 (如手机的电池、屏幕、相机等) 进行情感判别, 适合多属性产品评价。
- **多语言情感分析 (Multilingual Sentiment Analysis)**: 支持多种语言的情感分类, 面临跨语言、跨文化的挑战。

2.2 观点挖掘基本理论

观点挖掘 (Opinion Mining) 关注于发现文本中表达的观点目标、观点持有者及其态度。理论基础包括观点目标识别、观点表达抽取、观点极性判定等。

3 情感分析的基本流程

情感分析的典型流程包括:

1. 数据收集与预处理

- 收集需要分析的文本数据, 如客户评论、社交媒体帖子、新闻文章等。
- 预处理包括去除无关信息 (如 HTML 标签、特殊字符)、分词 (Tokenization)、去停用词 (如”的””了””and””the”等)、词干提取或词形还原 (Stemming/Lemmatization) 等。

2. 特征提取与表示

- 将文本转化为可用于建模的特征, 如词袋模型 (Bag-of-Words)、词嵌入 (Word2Vec、GloVe 等)。

3. 模型训练与分析

- 使用带标签的数据训练情感分类模型 (如正面、负面、中性), 常用方法包括规则法、机器学习、深度学习等。
- 训练完成后, 模型可对新数据进行情感预测。

4. 结果输出与应用

- 输出情感标签或情感分数, 应用于客户反馈分析、品牌监测、产品优化等实际场景。

(后续章节将详细介绍各类方法及其应用)

4 现有方法

4.1 词典方法

- **原理**：基于人工或自动构建的情感词典，将文本中的词与词典匹配，统计情感极性得分。
- **优点**：实现简单、可解释性强，适合无监督场景。
- **缺点**：依赖词典覆盖度，难以处理新词、隐喻、讽刺等复杂表达。

4.2 传统机器学习方法

- **流程**：
 1. **文本预处理**：分词（中文需特别注意，常用 jieba、HanLP 等）、去停用词、词性标注。
 2. **特征提取**：常用 TF-IDF、N-gram、词性、情感词统计等。
 3. **特征选择**：如卡方检验、互信息、信息增益等，减少高维稀疏问题。
 4. **分类模型**：SVM、朴素贝叶斯、逻辑回归、决策树等。
- **优缺点**：特征工程依赖人工经验，泛化能力有限。

4.3 深度学习方法

- **CNN**：适合提取局部上下文特征，常用于短文本情感分类。
- **RNN/LSTM/GRU**：适合建模长文本序列依赖，捕捉上下文信息。
- **注意力机制**：提升模型对关键信息的关注能力，常与 LSTM 结合。
- **多通道模型**：如多通道 CNN，将词向量、词性、位置等多种特征拼接输入，提升特征表达能力。

4.4 基于预训练模型的方法

- **BERT 及其变体**：通过大规模语料预训练，获得强大文本表示能力，微调后在情感分析、观点抽取等任务上表现优异。
- **ERNIE、RoBERTa**：进一步增强知识融合和语义理解能力，适合中文情感分析。
- **Fine-tuning 流程**：加载预训练模型，添加下游任务分类层，使用标注数据微调。

5 方法开发与评价

5.1 数据集

- **英文常用数据集**：IMDB、SST-2、Amazon Reviews、SemEval 等。
- **中文常用数据集**：ChnSentiCorp、NLPCC 微博情感分析、Douban 评论等。
- **数据预处理**：中文需分词，英文可直接用空格分词。注意去除噪声、表情符号、特殊字符等。

5.2 评估指标

- **准确率 (Accuracy)**: 整体分类正确率。
- **精确率、召回率、F1 值**: 适合不均衡数据集。
- **宏/微平均**: 多分类任务常用。
- **AUC、PR 曲线**: 用于模型综合性能评估。

6 总结与展望

6.1 研究总结

本文梳理了文本情感分析与观点挖掘的理论基础、主流方法及实验评价。深度学习和预训练模型极大提升了该领域的研究水平。

6.2 发展趋势

- 多模态情感分析 (结合文本、图像、音频等)
- 细粒度观点挖掘与解释性增强
- 跨语言、跨领域迁移学习

6.3 面临的挑战

- 数据标注成本高、领域适应性差
- 讽刺、隐喻等复杂情感表达难以识别
- 模型可解释性与公平性问题

7 工具与实践建议

7.1 常用工具

- **分词工具**: jieba、HanLP、LTP (中文); NLTK、spaCy (英文)
- **特征提取**: scikit-learn、gensim
- **深度学习框架**: TensorFlow、PyTorch
- **预训练模型库**: HuggingFace Transformers

7.2 典型流程

1. 数据采集与标注
2. 文本预处理 (分词、去噪、标准化)
3. 特征工程或词向量训练

- 4. 模型训练与调优
- 5. 结果评估与可视化

8 中文情感分析难点与前沿

- **分词歧义**：中文无天然分隔符，分词准确性影响后续分析。
- **情感表达复杂**：如反讽、双关、网络流行语等。
- **多粒度分析**：从句子、段落到篇章，甚至细粒度到观点目标级别。
- **多模态融合**：结合图像、音频等信息提升情感识别准确率。
- **可解释性与公平性**：深度模型”黑箱”问题，需提升可解释性。

9 典型应用案例与场景

9.1 社交媒体情感分析

案例：Nike 新品鞋款 Instagram 评论分析

Nike 作为全球知名运动品牌，在推出新款跑鞋时，利用情感分析技术对社交媒体（如 Instagram）上的用户评论进行分析。具体流程如下：

- **数据采集**：Nike 收集了过去一个月内所有与新品相关的 Instagram 帖子评论。
- **情感分类**：通过情感分析工具，将每条评论分为正面、负面或中性三类。
- **结果分析**：分析结果显示，60% 的评论为正面，30% 为中性，10% 为负面。正面评论主要集中在鞋子的设计、舒适度和性能上，负面评论则多涉及价格、尺码不合适和库存不足等问题。
- **决策优化**：Nike 根据分析结果，进一步强化了对产品设计和舒适度的宣传，同时针对负面反馈调整了定价策略和库存管理，提升了用户满意度和市场反响。

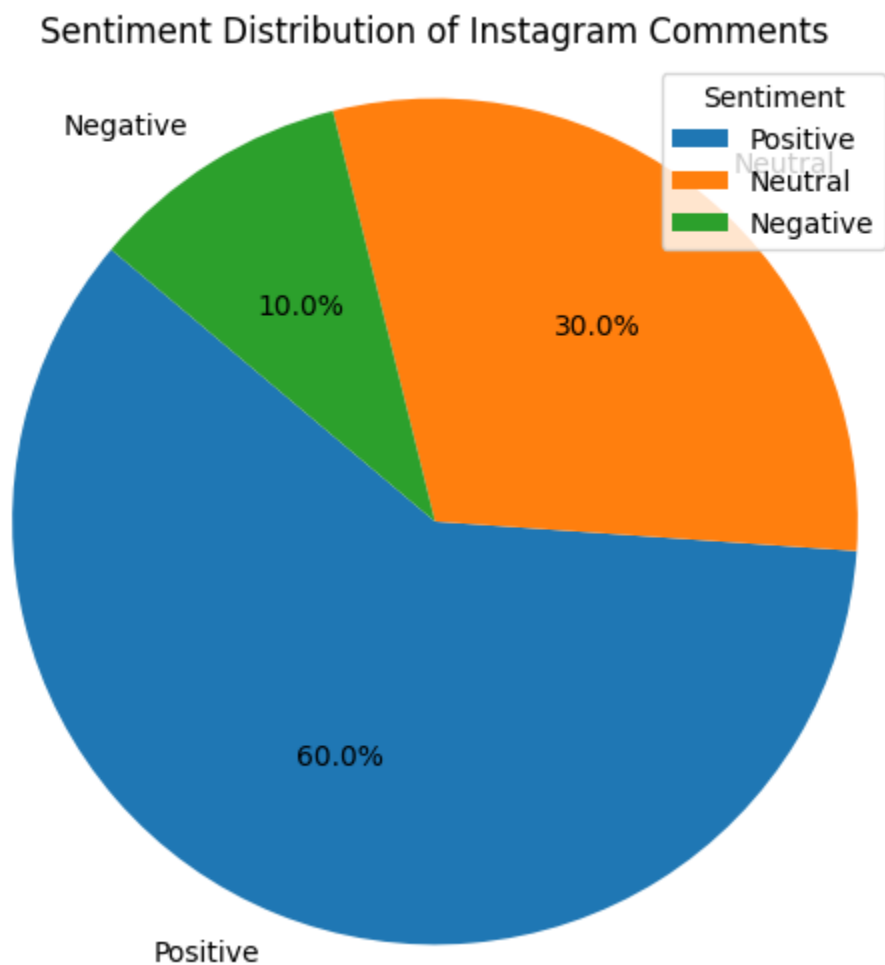


图 2: Nike 产品情感分析结果

该案例展示了情感分析在品牌营销和用户反馈收集中的实际价值，帮助企业及时把握用户情绪，优化产品和服务。

9.2 客户服务与产品改进

案例：Duolingo 在 Play Store 评论中的情感分析

Duolingo 是一款全球流行的语言学习应用。为提升用户体验，Duolingo 对 Google Play Store 上的用户评论进行了情感分析：

- **数据采集：**收集所有用户在特定时间段内的应用评论及评分（1-5 星）。
- **情感与评分关联分析：**将评论文本与评分结合，利用情感分析工具对文本内容进行正面、负面、中性分类。
- **发现问题：**分析发现，低分评论多与应用崩溃、课程难度大、内容单一等负面情绪相关，而高分评论则集中在界面友好、学习效果好、语言丰富等方面。
- **改进措施：**Duolingo 根据分析结果，重点修复了应用崩溃等技术问题，优化了课程内容和难度设置。后续监测显示，负面评论比例显著下降，平均评分提升，品牌形象和用户满意度也随之提高。

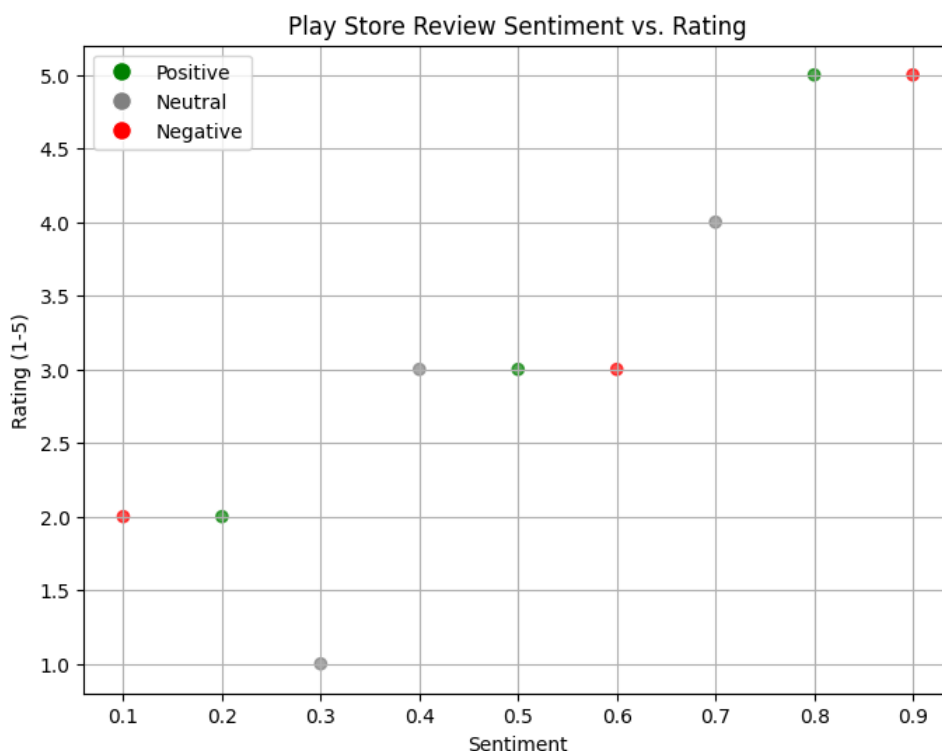


图 3: 客户评论情感分析

该案例说明情感分析不仅能帮助企业发现产品短板，还能指导产品优化和客户服务提升。

9.3 市场营销与新品推广

案例：护肤品公司新品市场情感分析

某护肤品公司在新品上市前，利用情感分析技术对市场反馈进行深度挖掘：

- **数据采集：**通过社交媒体、网络评论等渠道，收集与新品相关的用户评价和讨论。
- **情感分析：**利用情感分析工具对文本进行正面、负面、中性分类。
- **结果解读：**分析结果显示，70% 的评论为正面，20% 为中性，10% 为负面。正面评论多为天然成分、效果好、温和无刺激，负面评论则关注价格、包装和气味。
- **营销优化：**公司据此在后续宣传中突出产品的天然成分和温和特性，同时针对负面反馈调整了包装设计和定价策略，进一步提升了市场接受度和用户好评率。

该案例体现了情感分析在新品推广和市场调研中的重要作用，帮助企业精准把握用户需求，提升产品竞争力。

9.4 其他典型场景

- **客户反馈分析：**企业可通过情感分析了解客户对产品和服务的真实看法，及时改进，提升满意度。例如，电商平台可自动分析商品评论，发现用户关注的痛点和亮点。
- **品牌声誉管理：**实时监控社交媒体、评论平台等渠道的品牌舆情，及时应对负面信息，维护品牌形象。例如，企业可通过情感分析工具监测网络舆情，第一时间发现并处理危机公关事件。

- **竞争对手分析**：对比自身与竞争对手产品的情感分布，辅助战略决策。企业可分析竞争品牌的用户评论，了解其优势和不足，制定有针对性的市场策略。
- **营销活动效果评估**：通过分析活动期间的用户评论和讨论情感，评估营销效果，及时调整策略。例如，企业可在大型促销活动后，分析用户对活动的反馈，优化后续营销方案。

10 实践环节：基于 Python 的 Facebook 评论情感分析

10.1 实践环境与准备

- **操作系统**：Windows 10
- **Python 版本**：3.9
- **虚拟环境**：Anaconda/conda 本地目录环境（如 sentiment2）
- **依赖库**：nltk==3.7、pandas、matplotlib、numpy

10.1.1 环境搭建与依赖安装

```

1 # 1. 创建本地环境
2 conda create -p ./sentiment2 python=3.9 -y
3 conda activate ./sentiment2
4
5 # 2. 安装依赖
6 pip install nltk==3.7 pandas matplotlib numpy
7
8 # 3. 下载NLTK数据包
9 python
10 >>> import nltk
11 >>> nltk.download('punkt')
12 >>> nltk.download('wordnet')
13 >>> nltk.download('averaged_perceptron_tagger')
14 >>> nltk.download('vader_lexicon')
15 >>> exit()

```

10.1.2 数据准备

- 在实验目录下准备 kindle.txt，每行一条英文评论，内容涵盖正面、负面和中性情感。

10.1.3 脚本准备

- 在同一目录下准备 sentiment_analysis.py，内容如下：

```

1 import nltk
2 from nltk.tokenize import sent_tokenize, word_tokenize, PunktSentenceTokenizer
3 from nltk.stem.porter import PorterStemmer
4 from nltk.stem.wordnet import WordNetLemmatizer
5 from nltk.sentiment.vader import SentimentIntensityAnalyzer
6
7 with open('kindle.txt', encoding='ISO-8859-2') as f:
8     text = f.read()
9
10 # 分句与分词

```

```

11 tokenizer = PunktSentenceTokenizer(text)
12 sents = tokenizer.tokenize(text)
13 print('分词结果:', word_tokenize(text))
14 print('分句结果:', sent_tokenize(text))
15
16 # 词干化
17 porter_stemmer = PorterStemmer()
18 nltk_tokens = word_tokenize(text)
19 print('\n词干化:')
20 for w in nltk_tokens:
21     print("Actual: %s Stem: %s" % (w, porter_stemmer.stem(w)))
22
23 # 词形还原
24 wordnet_lemmatizer = WordNetLemmatizer()
25 print('\n词形还原:')
26 for w in nltk_tokens:
27     print("Actual: %s Lemma: %s" % (w, wordnet_lemmatizer.lemmatize(w)))
28
29 # 词性标注
30 print('\n词性标注:')
31 print(nltk.pos_tag(nltk_tokens))
32
33 # 情感分析
34 print('\n情感分析结果:')
35 sid = SentimentIntensityAnalyzer()
36 with open('kindle.txt', encoding='ISO-8859-2') as f:
37     for line in f.read().split('\n'):
38         if line.strip() == '':
39             continue
40         print(line)
41         scores = sid.polarity_scores(line)
42         for key in sorted(scores):
43             print('{0}: {1}, '.format(key, scores[key]), end='')
44         print()

```

10.2 运行与输出

在终端运行:

```
1 python sentiment_analysis.py
```

10.2.1 典型输出示例

```

1 分词结果: ['I', 'love', 'my', 'kindle', 'The', 'battery', 'life', 'is', 'amazing', ...]
2 分句结果: ['I love my kindle\nThe battery life is amazing\nThe screen is too small...']
3
4 词干化:
5 Actual: I Stem: i
6 Actual: love Stem: love
7 Actual: my Stem: my
8 Actual: kindle Stem: kindl
9 ... (略)
10
11 词形还原:
12 Actual: I Lemma: I
13 Actual: love Lemma: love
14 Actual: my Lemma: my

```

```

15 Actual: kindle Lemma: kindle
16 ... (略)
17
18 词性标注:
19 [('I', 'PRP'), ('love', 'VBP'), ('my', 'PRP$'), ('kindle', 'NN'), ...]
20
21 情感分析结果:
22 I love my kindle
23 compound: 0.6369, neg: 0.0, neu: 0.323, pos: 0.677,
24 The battery life is amazing
25 compound: 0.5859, neg: 0.0, neu: 0.513, pos: 0.487,
26 The screen is too small
27 compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
28 I am not satisfied with the customer service
29 compound: -0.3252, neg: 0.28, neu: 0.72, pos: 0.0,
30 ... (略)

```

```

Actual: clear Lemma: clear
Actual: I Lemma: I
Actual: enjoy Lemma: enjoy
Actual: reading Lemma: reading
Actual: every Lemma: every
Actual: day Lemma: day
Actual: It Lemma: It
Actual: is Lemma: is
Actual: just Lemma: just
Actual: okay Lemma: okay

词性标注:
[('I', 'PRP'), ('love', 'VBP'), ('my', 'PRP$'), ('kindle', 'NN'), ('The', 'DT'), ('battery', 'NN'), ('life', 'NN'), ('is', 'VBZ'), ('amazing', 'VBD'), ('The', 'DT'), ('screen', 'NN'), ('is', 'VBZ'), ('too', 'RB'), ('small', 'JJ'), ('I', 'PRP'), ('am', 'VBP'), ('not', 'RB'), ('satisfied', 'VBN'), ('with', 'IN'), ('the', 'DT'), ('customer', 'NN'), ('service', 'NN'), ('the', 'DT'), ('device', 'NN'), ('is', 'VBZ'), ('easy', 'JJ'), ('to', 'TO'), ('use', 'VB'), ('it', 'PRP'), ('is', 'VBZ'), ('a', 'DT'), ('bit', 'RB'), ('responsive', 'JJ'), ('reading', 'VBD'), ('on', 'IN'), ('a', 'DT'), ('manip', 'NN'), ('is', 'VBZ'), ('comfortable', 'JJ'), ('The', 'DT'), ('font', 'JJ'), ('size', 'NN'), ('is', 'VBZ'), ('adjustable', 'JJ'), ('Sometimes', 'RB'), ('it', 'PRP'), ('crashes', 'VBZ'), ('Delivery', 'NN'), ('was', 'VBD'), ('fast', 'RB'), ('The', 'DT'), ('d', 'NN'), ('is', 'VBZ'), ('readable', 'JJ'), ('clear', 'JJ'), ('cheap', 'JJ'), ('I', 'PRP'), ('enjoy', 'VBP'), ('it', 'PRP'), ('a', 'DT'), ('bit', 'RB'), ('more', 'RB'), ('expensive', 'JJ'), ('The', 'DT'), ('user', 'NN'), ('manual', 'NN'), ('is', 'VBZ'), ('not', 'RB'), ('clear', 'JJ'), ('I', 'PRP'), ('enjoy', 'VBP'), ('reading', 'VBD'), ('every', 'DT'), ('day', 'NN'), ('It', 'PRP'), ('is', 'VBZ'), ('just', 'RB'), ('okay', 'JJ')]

情感分析结果:
I love my kindle
compound: 0.6369, neg: 0.0, neu: 0.323, pos: 0.677,
The battery life is amazing
compound: 0.5859, neg: 0.0, neu: 0.513, pos: 0.487,
The screen is too small
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
I am not satisfied with the customer service
compound: -0.3252, neg: 0.28, neu: 0.72, pos: 0.0,
The device is easy to use
compound: 0.4859, neg: 0.0, neu: 0.633, pos: 0.367,
It is a bit expensive
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
Reading on kindle is comfortable
compound: 0.5149, neg: 0.0, neu: 0.548, pos: 0.452,
The font size is adjustable
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
Sometimes it crashes
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
Delivery was fast
compound: 0.0, neg: 0.0, neu: 1.0, pos: 0.0,
The design is simple and elegant
compound: 0.4767, neg: 0.0, neu: 0.617, pos: 0.383,
I wish it had more storage
compound: 0.4819, neg: 0.0, neu: 0.597, pos: 0.483,
The user manual is not clear
compound: -0.2924, neg: 0.384, neu: 0.696, pos: 0.0,
I enjoy reading every day
compound: 0.6059, neg: 0.0, neu: 0.684, pos: 0.516,
It is just okay
compound: 0.2243, neg: 0.0, neu: 0.612, pos: 0.388,

```

图 4: 终端运行结果

10.3 过程总结与经验

- NLTK 3.8 及以上版本在部分环境下会出现 `punk_tab`、`averaged_perceptron_tagger_eng` 等异常，建议降级至 3.7。
- NLTK 数据包建议用标准包名下载（如 `averaged_perceptron_tagger`），不要用 `_eng` 结尾。
- 保证脚本、数据和环境在同一目录，避免路径和权限问题。
- 典型输出包括分词、分句、词干化、词形还原、词性标注和情感分析分数，便于后续分析和可视化。

10.4 参考与延伸

- GeeksforGeeks: Facebook Sentiment Analysis using python
- 可尝试将流程应用于中文评论，或结合深度学习方法进一步提升分析效果。

10.5 本次实验方法分析

本次实验采用的是词典方法中的 VADER 情感分析器：

- **VADER (Valence Aware Dictionary and sEntiment Reasoner)** 是基于情感词典和规则的英文情感分析工具。
- 它通过内置的情感词典和极性规则，对每条评论进行正面、负面、中性和综合 (compound) 得分的打分。
- 这种方法不需要训练数据，属于**无监督的词典方法**，适合社交媒体、短文本等场景。

10.5.1 方法优缺点分析

优点：

- 实现简单，开箱即用，无需训练。
- 对英文社交媒体文本、短句、表情符号等有较好适配。
- 结果可解释性强，输出正负面比例和综合分数。

缺点：

- 依赖内置词典，遇到新词、俚语、讽刺、复杂语境时准确率有限。
- 对中文、长文本、专业领域文本适应性较差。
- 无法自动学习新情感表达，泛化能力有限。

参考文献

- [1] Liu B. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
- [2] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL, 2019.
- [3] Sun Y, Wang S, Li Y, et al. ERNIE: Enhanced Representation through Knowledge Integration. arXiv preprint arXiv:1904.09223, 2019.
- [4] Pang B, Lee L. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2008.
- [5] 王斌, 刘挺. 中文情感分析研究综述. 中文信息学报, 2014.
- [6] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP, 2002.
- [7] Minqing Hu and Bing Liu. Mining and summarizing customer reviews. KDD, 2004.
- [8] Maite Taboada, et al. Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, 2011.

- [9] Duyu Tang, Furu Wei, Bing Qin, Ting Liu, Ming Zhou. Building Large-Scale Twitter-Specific Sentiment Lexicon: A Representation Learning Approach. COLING, 2014.
- [10] Zichao Yang, et al. Hierarchical Attention Networks for Document Classification. NAACL, 2016.