

实验报告：多模态情感分析

10225501463王江扬

仓库地址

传送门

具体实现

在代码中已经详细备注

1. 实现问题与解决方案

数据加载与预处理

- 问题：** 文件路径错误、分词器填充不一致。
- 解决：** 添加路径检查，设置 `padding='max_length'` 保证张量一致。

训练效率

- 问题：** ResNet-50 训练中 GPU 内存不足。
- 解决：** 调整批量大小，使用梯度累积优化内存利用。

损失函数错误

- 问题：** 标签格式不匹配。
- 解决：** 确保标签格式转换为整数。

2. 模型设计

特征提取

- 文本：基于 BERT 提取语义特征。
- 图像：通过 ResNet-50 提取视觉特征。

融合策略

- 使用特征拼接融合文本与图像，配合全连接层捕获模态间的交互。

分类结构

- 单层线性分类头，输出三分类结果（正面、中性、负面）。

3. 模型亮点

- 多模态集成：**文本与图像互补，提升情感分析精度。
- 预训练优势：**利用 BERT 和 ResNet-50 的预训练权重，高效提取特征。
- 轻量级设计：**融合策略简单，易于扩展和部署。
- 结果稳定性：**消融实验表明多模态融合具有显著优势，验证了模型的稳健性。

4. 验证集性能

多模态模型

- 准确率：** 85.7%
- F1 分数：** 86.2%

消融实验

模型配置	准确率	F1 分数
仅文本 (BERT)	81.3%	82.1%
仅图像 (ResNet)	78.5%	79.8%
文本+图像	85.7%	86.2%

5. 观察与改进

- 多模态优势：**模型融合明显优于单模态基线。
- 文本贡献：**文本特征在情感分类中占主导地位。
- 图像作用：**虽然图像特征贡献较小，但在细粒度情感判断中仍提供增益。
- 改进方向：**
 - 引入注意力机制优化模态交互。
 - 增强图像数据扩展模型泛化能力。
 - 使用更大数据集提升鲁棒性。
 - 探索其他融合策略（如门控机制）进一步提升性能。