

当代人工智能实验五 - 多模态情感分析任务

1. 实验概述

1.1 实验内容

设计和实现一个多模态融合模型，用于分析包含文本和图像的社交媒体内容的情感倾向。具体来说，模型需要对输入的文本-图像对进行处理，并将情感分类为积极（positive）、中性（neutral）和消极（negative）三类。

1.2 数据集介绍

能够观察到，实验使用的数据集似乎是一个匿名的社交媒体数据集，其中每个样本包含一段文本内容和一张配图，并通过唯一的guid进行标识。

- 文本内容包含多种语言，并带有典型的社交媒体特征，如 hashtag 标签和 @提及
- 图像内容与文本相关，共同表达用户的情感倾向

示例：

RT @SimpsonsQ0TD: "Oh, don't you worry, most of you will never fall in love and marry out of fear of dying alone."

[标签: negative]

2. 实验设计

2.1 数据处理 (utils/data_preprocessing.py)

- 文本：使用 BERT tokenizer 进行分词，并设置最大序列长度为 128，对超长文本进行截断，对较短文本进行填充。考虑到数据集中存在多语言文本，我们选择了支持多语言的预训练模型。
- 图像：
 - 基础预处理：统一调整图像尺寸为 224×224，并使用 ImageNet 数据集的均值和标准差进行归一化
 - 数据增强：
 - 随机水平翻转：增加模型对图像水平方向变化的鲁棒性

2. 随机旋转 (± 15 度)：提高模型对角度变化的适应能力
3. 颜色封闭：调整亮度、对比度和饱和度，范围均为 ± 0.2

```
transforms.Resize((224, 224)),  
# 以下为数据增强部分，若不需要则注释  
transforms.RandomHorizontalFlip(),      # 随机水平翻转  
transforms.RandomRotation(15),          # 随机旋转  
transforms.ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2), # 颜色封闭
```

- 数据集划分：按照 7:3 的比例将训练数据集划分为训练集和验证集，并使用固定的随机种子（SEED=42）确保实验的可复现性。

2.2 模型设计

2.2.1 文本特征提取器

文本特征提取模块采用了基于 Transformer 的预训练模型架构，支持 3 种 backbone 选择：

1. BERT：使用 bert-base-uncased 作为基础模型
2. DistilBERT：作为 BERT 的轻量化版本，用于提升训练效率
3. RoBERTa