**华东师范大学数据科学与工程学院上机实践报告**

|  |  |  |
| --- | --- | --- |
| **课程名称**：当代人工智能 | **年级**：2022级 | **上机实践成绩**： |
| **指导教师**： | **姓名**：曾玮 |  |
| **上机实践名称**： | **学号**：10225501412 | **上机实践日期**：2025.1.18 |
| **上机实践编号**： | **组号**： |  |

**设计+亮点**

1. 多模态融合：

该模型结合了文本和图像两种模态的信息。文本信息通过BERT提取，图像信息通过ResNet50提取。通过融合这两种特征，模型可以更全面地理解输入数据。

2. 使用预训练模型：

使用BERT和ResNet50作为特征提取器。这些模型在大规模数据集上预训练过，能够提供强大的特征表示能力，从而提高下游任务的性能。

3. 灵活的特征融合：

通过简单的线性层将文本和图像特征进行融合，并使用ReLU激活函数和Dropout层来增强模型的非线性表达能力和防止过拟合。

4. 轻量级文本模型：

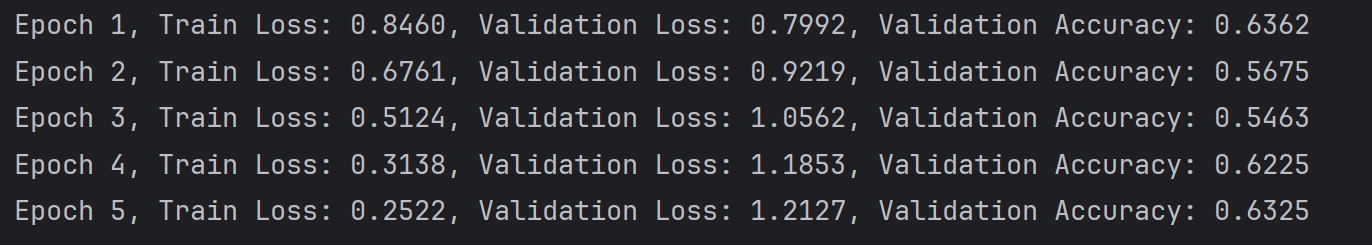
BERT具有更少的参数和更快的推理速度，同时在许多任务上仍能保持良好的性能。这使得模型在资源受限的环境中也能高效运行。

超参数选择：

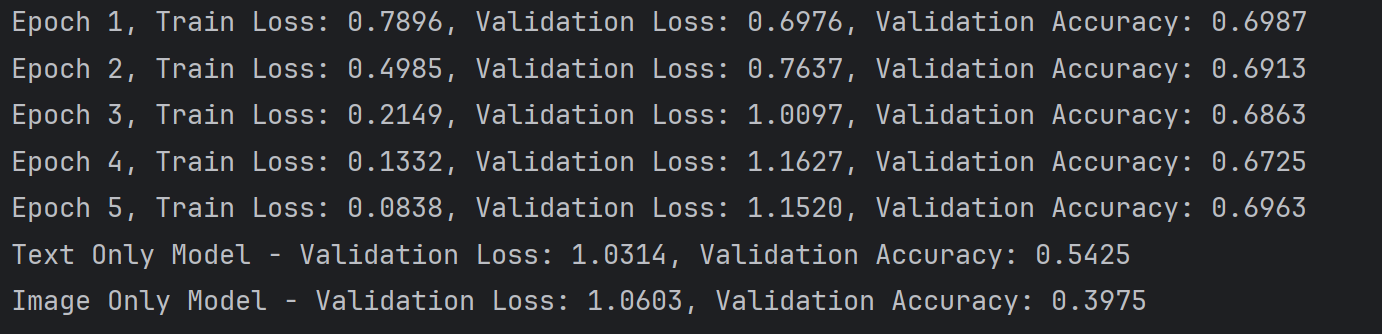
学习率：5e5 epoch：5

Max Sequence Length：128 dropout：0.3

结果 ：



结果 + 消融实验：



结果保存在predications中

多模态融合模型在验证集上的表现优于单独使用文本或图像的数据。这说明文本和图像的结合能够提供更丰富的信息：

1. 多模态模型：

验证准确率在0.6725到0.6987之间，表现相对稳定。

训练损失和验证损失逐渐下降，说明模型在学习过程中有效地捕捉到了数据的特征。

2. 文本模型：

验证准确率为0.5425，低于多模态模型。这表明仅使用文本信息时，模型可能无法捕捉到某些需要图像信息才能理解的特征。

3. 图像模型：

验证准确率为0.3975，明显低于多模态模型。这可能是因为文本信息在情感分析中提供了重要的上下文，而仅靠图像难以准确判断情感，也是本模型的缺陷所在，无法完整的从图像中获取信息。

**BUG：**

编码问题：发现data文件夹中txt文件编码格式不统一，有utf8，gb2312等，所以采用了chardet 自动检测文件编码

