

作业 6

郭中贺

2022 年 6 月 5 日

---

## 调研报告部分

# 1 关于媒体与认知技术在新冠肺炎诊断方面应用的调研报告

## 1.1 问题背景

自新冠疫情爆发以来，人们的生活与工作受到了很大影响，在医疗科学领域，如何精确而正确地诊断新冠肺炎是一个重要问题。唐智贤 等 (2022) 指出，到目前为止，X 射线扫描和 CT 为诊断肺部疾病的首要检查方法，然而肺部影像不仅有着复杂的表现，而且目前我国肺部影像的数据量逐年增长，多种因素导致影像科医生任务繁重。因此，为了解决这种医疗问题，基于人工智能的疾病诊断技术应运而生，为医学影像检查提供了广阔应用前景。

## 1.2 已有措施与技术

He et al. (2021) 研究人员，一方面通过肺叶分割和联合执行严重度评估，进而对原始三维 CT 图像中的新冠肺炎严重度进行自动评估。另一方面，他们开发了一个多任务、多实例 M2UNET 网络评估新冠肺炎患者的严重程度。

徐翠莲 等 (2021) 研究人员利用基于深度学习的 AI 模型实现肺炎病灶的自动分割，从每一帧图像的病灶区域中提取影像组学特征，经过 LASSO 回归降维后分别采用高斯朴素贝叶斯、随机森林和极端梯度提升的方法在训练集中建立影像组学模型，并在验证集中测试模型性能。采用 Dice 系数评估肺炎病灶分割的准确性，采用 ROC 评估新冠肺炎诊断效能。

## 1.3 技术效果分析

在肺炎病灶分割方面，需要以人工标记结果为标签，利用图像处理方法以及深度学习来实现对 CT 影像中验证病灶的自动分割，研究人员的结果也表明，AI 分割的病灶与人工标记结果有着非常好的一致性。

对于提取的影像组学特征，我们可以使用一些方法对特征进行降维，保留关键信息，去除冗余信息。

之后，我们可以基于提取的特征，来对 CT 图像进行新冠肺炎的结果诊断学习，徐翠莲等研究人员的实验也证明，基于高斯朴素贝叶斯的模型获得了相对较好的效能，其模型在薄层 CT 影像验证集上的 AUC 分别为 0.919、0.838 和 0.829，在厚层 CT 影像验证集上的 AUC 分别为 0.802、0.730 和 0.715。

## 1.4 未来优化思路

虽然目前人工智能在上述医疗方面发挥了重要的作用，但是由于技术原因和其他因素的制约，还存在一定的问题。例如，基于机器学习的方法对医学影像进行诊断必定要基于大量的数据，否则学习效果不能使人满意。而这些医学影像作为数据输入时，通常需要提前进行人工标注。因此，在未来的工作中，我们需要建立更加完善的医学影像数据库，从而有效提升图像的临床代表性、标注规范性与图像多样性。另一方面，我们还可以进一步开发半监督甚至无监督的学习模型来对医学影像进行诊断，提高算法的泛用性。

## 编程部分

## 2 编程作业报告

### 2.1 代码补全

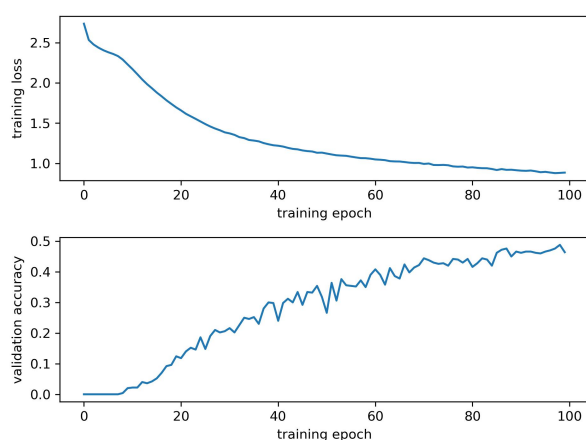
补全代码后，运行命令 `python network.py`，以验证代码的正确性。结果如下：

```
(myenv) C:\Users\siwen\media-and-cognition\hw6>python network.py
The output size of model is correct!
```

如图，可以看出网络输出变量的尺寸是正确的，可以看出 Transformer 层实现完成。

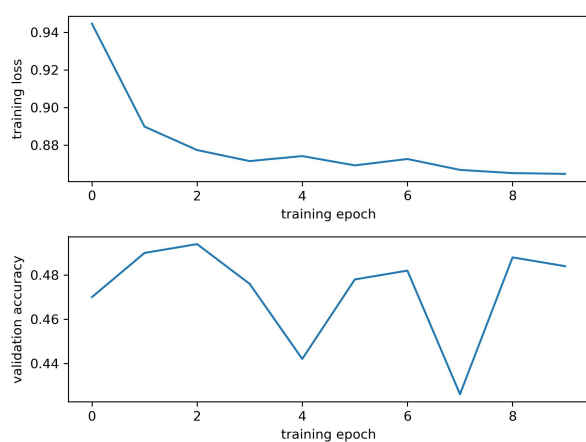
### 2.2 模型训练

输入如下命令：`python main.py -mode train` 进行训练 100 轮，得到训练结果如下：



可以看到，训练完成 100 轮后，模型在验证集上的 loss 降低到 0.8 左右，准确率较低，但也达到了 50% 左右。

使用命令 `python main.py -mode train -load_pretrain -pretrain_path models/model_epoch100.pth -epoch 10` 将模型继续训练 10 轮之后，得到结果如下：



其中最后一轮的 loss 和 accuracy 分别为 0.865 和 48.4%：

```
Epoch [10/10] start ...  
train loss = 0.865, validation word accuracy = 48.4%
```

由于我们使用的数据量较小所导致的最终在验证集上的正确率大概只有 50% 左右，从 loss 和 accuracy 变化曲线中我们可以发现，训练初期由于模型在学习过程中，初期的 loss 下降较慢，准确率也一直为 0，而随着训练的逐渐进行，大概从第 10 轮开始，准确率开始逐渐上升，到 80 轮之

后, loss 下降速度减缓, 准确率上下波动开始较为明显。总的来说, 训练的结果还是较为理想的。可以看出, 与上次作业中采用 RNN 网络模型相比, 该次训练过程要更加困难, 这与 transformer 本身特性有关。

**transformer 在训练时具有并行训练的特性, 原因在于其自注意的机制, 在解码时每个时刻的输入依赖的是正确的样本, 而非上一时刻的输出。transformer 实现 Decoder 部分训练并行化, 就是一次性将整个目标句子输入给 decoder, 然后利用自注意力算法计算出结果, 往后面网络继续传递, 最后计算出预测值, 分别对应每个时刻的输出, 从而实现并行训练。**

## 2.3 使用训练好的模型预测新的文本图像

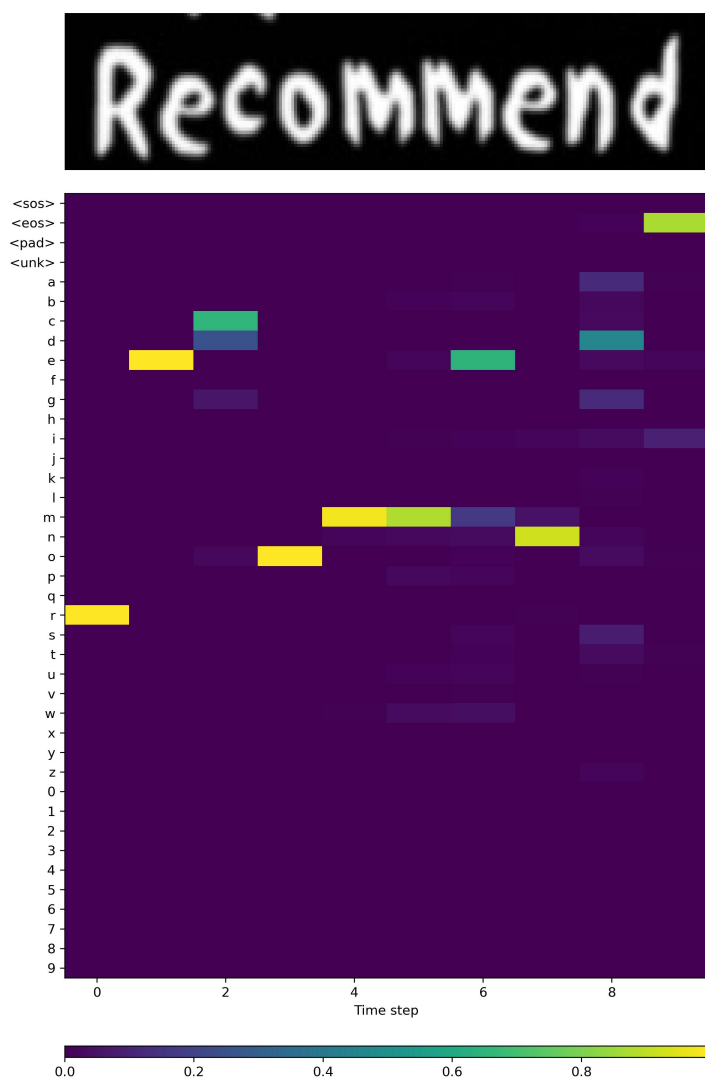
使用命令 `python main.py --mode predict --in_path data/my_own/recommend.png`, 预测新的文本图像。输入图像为:



使用上面一共训练 110 轮的模型进行预测。输出结果如下:

```
(myenv) C:\Users\siwen\media-and-cognition\hw6>python main.py --mode predict --in_path data/my_own/recommend.png --model_path models/model_epoch10.pth
[Info] Load model from models/model_epoch10.pth
prediction: recommend
Transformer visualization has been saved as data/my_own/recommend_vis.jpg
```

可视化结果如下:



可以看到得到了正确的预测结果 recommend，结合可视化结果我们可以看到这里预测的结果为 recommend-., 能够正确识别，一方面是由于输入图像较为理想，字符清晰易分辨，另一方面也可以看出模型有一定的泛用性。上一次的作业的可视化结果中，我们发现不同字母之间通常会间隔一个或多个空白符，通过 CTC 解码可以得到正确的结果。而本次作业使用的 transformer 解码，能够直接输出正确字符序列。

**更进一步地说，transformer 解码和 CTC 解码的区别在于：**

Transformer 是属于 Attention-based 的模型，通过注意力机制对各个输入进行联合建模需要一次获取整个输入语音编码序列。在 Decoder 的时候，

是需要根据之前的翻译，求解当前最有可能的翻译，CTC 则是基于条件独立性假设，每个时刻的输出仅依赖于当前时刻的输入，不需要等待完整的上文输入，通过引入空白符的方式，保持输入与输出的单调对齐，使用动态规划的算法实现快速的解码。

## 2.4 本次作业遇到的问题及解决方法

无。

## 2.5 对本次作业的意见及建议

本次作业也较为简单，通过代码注释中的一步一步教程，我们能很快了解下一步该做什么，该如何实现，感谢老师和助教的悉心指导

## 参考文献

唐智贤, 王一森, 周靓怡, 等. 人工智能技术在肺部影像辅助诊断中的应用进展[J]. 中国医学物理学杂志, 2022, 39: 655-660.

He K, Zhao W, Xie X, et al. Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of covid-19 in ct images[J]. Pattern Recognition, 2021, 113: 107828.

徐翠莲, 印宏坤, 卮立荣, 等. 深度学习结合影像组学的胸部 CT 新冠肺炎智能诊断模型研究[J]. 罕少疾病杂志, 2021, 28: 16-20.