

基于深度学习的小样本医学影像分析方法研究

全昱

(延边大学 经济管理学院 吉林 延吉 133002)

[摘要] 医学影像数据具有处理繁琐, 专业性强等特点, 因此要获取大量的可以用于模型训练的医学图像数据具有一定的难度。本文从小样本的情况出发, 进行探究。在样本数量较少的时候, 通过对深度学习网络模型的改进, 通过相关的集成学习方法研究, 解决普遍存在的医学影像不足的问题。针对小样本影像数据, 探索更精确的医学检测方法, 从而辅助临床医生做出更合理的决策, 进而减少病患不必要的活体检查, 提升其生存概率及生活质量。

[关键词] 医学影像; 集成学习; Adaboost算法

[DOI] 10.12252/j.issn.2096-6288.2021.06.779

引言

近年来, 计算机视觉与人工智能快速发展, 医学影像和计算机辅助诊断成为了医学和计算机视觉的新兴交叉学科^[1]。医学影像数据具有庞大、客观、可量化的特点, 符合深度学习的应用场景^[6], 也因此促使了人工智能与医学领域的双向共同发展。医学成像技术的不断突破, 推动了生命科学的革命。将人工智能用于医疗影像分析, 可以帮助医生更精准的对病情进行分析诊断。目前, 有超过90%的医疗数据来自医疗影像, 将这些数据运用人工智能算法自动分析影像, 结合病人的病历记录综合对比分析, 就可以大大降低医学误诊率^[2]。

但是众所周知, 网络模型的训练与参数的优化需要大量的有标注的数据来完成, 而医学影像的标注不仅耗时耗力, 并且对专业性要求极高, 需要具有丰富经验的专业人士进行标注, 这就加大了数据收集的难度。本文从小样本的情况出发, 进行探究。在样本数量较少的时候, 通过对深度学习网络模型的改进, 通过相关的集成学习方法研究, 解决普遍存在的医学影像不足的问题^[4]。本文提出一个基于Adaboost算法的双特征输入法, 用于医学影像的特征处理问题, 将图像的纹理特征与病变特征进行并行训练, 在输出强分类之前, 可以通过两种分类结果进行反馈调节, 进一步更新样本权重。

1. 单输入的Adaboost分类算法

Adaboost算法是一种集成算法^[3], 可以将几个弱分类器集成在一起组成强分类器。Adaboost模型将几个弱分类器通过多次迭代训练弱分类器来预测样本的标签, 根据预测结果按照样本权重调节机制进行相应调节, 当迭代次数达到一定值时, 根据之前的训练结果, 将各个弱分类器与其分类权重进行线性加权组成强分类器。

单输入的Adaboost分类算法的迭代过程为, 首先将串联合并的纹理特征与病变特征合并成一个特征输入到第一个弱分类器中, 通过训练可以得到三个量: 分类误差、分类器权重和更新后的样本权重, 其中, 被错误分类的样本权重会增加, 反之正确的样本权重会减小^[5]; 接下来, 第二个弱分类器进行迭代, 与上述过程相同, 会得到三个量的更新值; 同样的, 其余

的弱分类器按此规定依次进行迭代; 最后, Adaboost算法会对其进行一个叠加操作。其中, 算法公式如下所示:

(1) 训练样本权重: 将每个N样本权重等分, D_1 表示权重集合。公式为:

$$D_1 = \left(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \right) \quad (1.1)$$

(3) 计算分类器误差: 分类精度越高, 分类器的权重越大。公式为:

$$\varepsilon_t = \sum_{i=1}^N D_i^t [h_t(x_i) \neq y_i] \quad (1.2)$$

其中, 其中 $h_t(x_i)$ 表示第t次迭代中样本 x_i 的分类类别, y_i 表示样本 x_i 的真实标签。

(3) 计算分类器权重: 根据分类误差确定弱分类器在最终强分类器中的比

例系数。公式为:

$$\alpha_t = \frac{1}{2} \lg \frac{1 - \varepsilon_t}{\varepsilon_t} \quad (1.3)$$

更新样本权重: 根据分类结果, 调整样本权重。公式为:

$$D_i^{t+1} = D_i^t \cdot \exp(\alpha_t) [h_t(x_i) \neq y_i] \quad (1.4)$$

输出强分类器: 将每个弱分类器权重与对应的分类精度进行线性叠加

构成强分类器, 公式为:

$$H(x) = \sum_{t=1}^T \alpha(t) [h_t(x_i) = y_i] \quad (1.5)$$

2. 基于Adaboost的双特征输入算法

在基于Adaboost的双特征输入算法中, 与单输入算法的区别在于: 单输入模型的输入是将纹理特征与病变特征通过串联的方式进行输入, 而双特征输入的模型训练中是将两种特征并联输入。另外, 单输入模型每个弱分类器迭代的输出构成了强分类器, 而在双特征输入模型中, 每次迭代后, 两种特征分类器进行相互反馈, 以调节错误分类的样本。

其中, 两个分类器所占的权重可分别表示为:

$$\alpha_t^i = \frac{1}{2} \lg \frac{1 - \varepsilon_t^i}{\varepsilon_t^i} \quad (2.1)$$

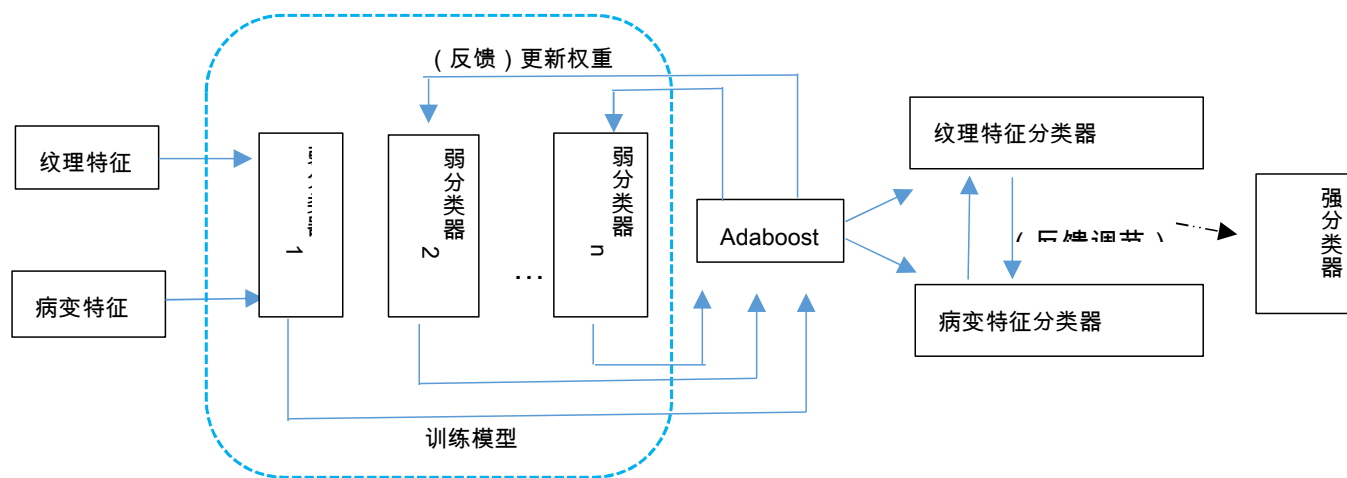


图2-1 基于Adaboost的双特征输入算法

$$\beta_i = \frac{1}{2} \lg \frac{1 - \varepsilon_2^i}{\varepsilon_2^i} \quad (2.2)$$

新的样本权重可表示为:

$$D_i^{t+1} = D_i^t \cdot \exp\left(\frac{1}{2} \alpha_i \cdot [h(x_i) \neq y_i] + \frac{1}{2} \beta_i \cdot [h(x_i) \neq y_i]\right) \quad (2.3)$$

最终的强分类器可表示为:

$$H(x_i, y_i) = \sum_{t=1}^T (\alpha(t)[h(x_i) = y_i] + \beta(t)[h(x_i) \neq y_i]) \quad (2.4)$$

迭代过程如图2-1所示:

如图, 首先将两种特征并行输入到第一个弱分类器中进行独立训练, 迭代结束后会分别得到两种特征分类器的分类误差与分类器权重, 然后根据两个分类器的分类结果进行相互反馈, 调节错误分类的样本。接下来, 按照同样的规则进行迭代操作, 最后, 将两种分类器的分类精度与分类权重各自相乘, 并将结果相加构成最终的强分类器。

结语

在小样本的情况下, 对于神经网络模型的训练, 会因为数据的不足而导致模型训练不完善, 从而使得模型处理数据的准确率降低; 同时, 某些情况下, 会因为数据太少而出现过度拟合的情况。因此, 提出将集成学习的方法用于深度学习神经网络模型在医学影像分析方面的训练, 以提高影像分析的准确率。利用以上方法可以提高样本特征分类的准确率, 在有限的样本数据下, 得到更为精准的影响分析结果。希望本文所提出的观点可以为相关研究提供有效思路, 可以对相关领域的研究有所

帮助。

参考文献

- [1] Papademetris X, DeLorenzo C, Flossmann S, et al. From medical image computing to computer-aided intervention: Development of a research interface for image-guided navigation[J]. The International Journal of Medical Robotics and Computer Assisted Surgery, 2009, 5(2): 147-157.
 - [2] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436.
 - [3] Dubout C, Fleuret F. Adaptive sampling for large scale boosting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1431-1445.
 - [4] 何雪英, 韩忠义, 魏本征. 基于深度学习的乳腺癌病理图像自动分类[J]. 计算机工程与应用, 2018, 54(12): 121-125. DOI: 10.3778/j.issn.1002-8331.1701-0392.
 - [5] 徐日, 张溢. 基于Adaboost算法的推荐系统评分预测框架[J]. 计算机系统应用, 2017, 26(8): 107-113.
 - [6] 张西宁, 刘书语, 余迪, 等. 改进深度卷积神经网络及其在变工况滚动轴承故障诊断中的应用[J]. 西安交通大学学报, 2021, 55(6): 1-8.
- (基金项目: 延边大学应用基础项目“基于深度学习的医学影像分析在精准医疗中的应用”)