

几种机器学习方法在黑色素瘤计算机 辅助诊断中的性能比较*

王 婷¹, 张 宁¹, 后桂荣², 余学飞^{1†}

(1. 南方医科大学 生物医学工程学院, 广州 510515; 2. 南方医院 皮肤科, 广州 510515)

摘 要: 黑色素瘤的计算机辅助诊断是基于激光共聚焦扫描显微镜(CLSM)皮肤图像纹理特征,并引入机器学习的技术,为临床应用研发的一种能够准确、有效地识别在体恶性黑色素瘤新医学诊断方法,将常用的基于机器学习的ID3、分类与回归树(CART)和AdaBoost三种算法应用于良恶性黑色素瘤图像的特征识别,并对各种学习方法的性能进行比较。实验结果表明,AdaBoost算法具有较好的分类识别性能,不但提高了恶性黑色素瘤早期诊断的准确度,降低了良性黑色素瘤的误诊率,而且为临床上早期发现和诊断提供了客观依据。

关键词: 黑色素瘤; 计算机辅助诊断; 机器学习

中图分类号: TP301

文献标志码: A

文章编号: 1001-3695(2013)06-1731-03

doi:10.3969/j.issn.1001-3695.2013.06.034

Performance comparison of several machine learning methods for computer-aided diagnosis of melanoma

WANG Ting¹, ZHANG Ning¹, HOU Gui-rong², YU Xue-fei^{1†}

(1. School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China; 2. Dep. of Dermatology, Nanfang Hospital, Guangzhou 510515, China)

Abstract: Computer-aided diagnosis of melanoma is a new medical diagnosis method, based on the texture features of confocal laser scanning microscopy (CLSM) skin images, it uses the machine learning methods, for the accuracy of diagnosis of malignant melanoma and valid identification of malignant melanoma. The paper applied currently popular machine learning methods including the ID3 algorithm, classification and regression tree (CART) algorithm, AdaBoost algorithm, to features detection between benign and malignant melanoma images. And then it compared the performance of each learning method. Experimental results show that the AdaBoost algorithm has the best discrimination performance. It improves the early malignant melanoma diagnosis accuracy, and reduces the misdiagnosis rate of benign common nevi, also it provides an objective basis for clinical diagnosis and early detection.

Key words: melanoma; computer aided diagnosis; machine learning

0 引言

激光共聚焦扫描显微镜(CLSM)由于其具有的超高分辨率、良好图像对比度和连续光学切片等功能,被作为一种新型的用于恶性黑色素瘤早期发现与诊断的理想辅助成像工具^[1,2]。计算机辅助医学诊断是指通过计算机模拟专家思维和推理过程,对病情进行判断,得出医学专家诊断和治疗方案。本文黑色素瘤的计算机辅助诊断主要是基于 CLSM 皮肤图像,并结合计算机的分析计算,辅助医师发现病灶,提高了临床诊断的准确率、敏感性和特异性^[3]。

黑色素瘤与痣在外观与特征上有很高的相似性,且大部分由痣转变而来,因此正确区分黑色素瘤与良性痣仍是临床医学诊断中的一个难题。本文将依据 200 例常见良性痣与 200 例黑色素瘤 CLSM 图像,提取基于小波变换的纹理特征,并利用机器学习中的相关理论与算法实现在体恶性黑色素瘤的计

机辅助识别^[4]。决策树方法中 ID3 算法、CART 算法和集成学习方法中的 AdaBoost 算法作为当前主流的机器学习算法的代表,在许多学科领域中都有应用,尤其在模式识别领域中更是得到广泛认可。本文将这几种机器学习方法分别应用于良恶性黑色素瘤图像特征的识别,并在 MATLAB 开发环境下对各学习方法进行实验。通过比较发现 AdaBoost 算法具有较好的分类识别性能,为临床上早期发现和诊断黑色素瘤提供了参考和依据。

1 基本原理

1.1 黑色素瘤的计算机辅助诊断原理

黑色素瘤和痣的 CLSM 图像在不同尺度上的建筑结构对 CLSM 图像诊断起着重要作用,因此小波变换因其频谱特性而适合于图像的自动分析识别。计算机辅助诊断原理主要是

收稿日期: 2012-09-11; 修回日期: 2012-11-05 基金项目: 广东省科技计划项目(2011B031800087); 广州市科技计划项目(2010J-E361)

作者简介: 王婷(1987-),女,湖北黄冈人,硕士,主要研究方向为医学图像处理与智能医学仪器;张宁(1978-),男,河南洛阳人,讲师,博士,主要研究方向为医学图像分析;后桂荣(1980-),女,河南濮阳人,主治医师,硕士研究生,主要研究方向为皮肤影像学;余学飞(1965-),男(通信作者),湖北浠水人,教授,博士,主要研究方向为医疗仪器设计与医学图像分析(xuefeiyu@fimmu.com)。

基于 CLSM 图像,通过小波变换将所有图像分解成不同尺度的频率带信息,而这种信息对于信号分类是非常有用的。因为图像数据是非结构化数据,不能直接用于分类。图像的小波变换可用于图像特征提取,实际上是将小波变换看做一种特征映射。提取的两类图像纹理特征经由特征评价分析具有很好的区分度,因此可作为下一步的分类特征。黑色素瘤的计算机辅助诊断主要任务就是一个分类识别的过程,基于以上两类图像的纹理特征,利用机器学习相关方法判断是黑色素瘤图像还是良性痣图像^[5-7]。其原理流程如图 1 所示。

1.2 几种机器学习方法的工作原理

机器学习是计算机科学中研究怎么让机器具有学习能力的分支。目前,机器学习方法已经在科学研究、语音识别、人脸识别、手写识别、游戏等领域中得到应用。本文将机器学习方法应用于医疗诊断中,并具体介绍决策树学习方法和集成学习方法^[8,9]两类经典热门的方法。

1.2.1 分类与回归树

Breiman 等人提出一种基于决策树的统计模型,即分类和回归树(classification and regression trees, CART)算法。它是一种具有纵向分析特征树型结构,其分析结果的过程与临床思维十分相似,因而容易被临床医生理解和接受。其基本原理是通过对由测试变量和目标变量构成的训练数据集的循环分析而形成二叉树形式的决策树结构^[10]。CART 算法采用经济学中的基尼系数(Gini index)作为选择最佳测试变量的准则。基尼系数的定义如下:

$$\text{Gini index} = 1 - \sum_j P^2(j/h) \quad (1)$$

$$P(j/h) = \frac{n_j(h)}{n(h)}, \sum_j P(j/h) = 1 \quad (2)$$

其中: $P(j/h)$ 是从训练样本集中随机抽取一个样本,当某一测试变量值为 h 时属于第 J 类的概率; $n_j(h)$ 为训练样本中测试变量值为 h 时属于第 J 类的样本个数; $n(h)$ 为训练样本中该测试变量值为 h 的样本个数; J 为类别个数。

CART 算法是从众多的预测属性(模型的输入属性)中选择一个属性或多个属性的组合,作为树节点的分裂变量,把测试变量分到各个分枝中,重复该过程建立一棵充分大的分类树。但按照上述过程生成的完整决策树往往会出现过度拟合现象,因此 CART 算法采用交叉验证的方法进行剪枝,剪枝算法使用独立于训练样本集的测试样本集对子树的分类错误进行计算,找出分类错误最小的子树作为最终的分类模型。

1.2.2 ID3

Quinlan 提出的 ID3 算法是最早有影响的决策树算法,区别于 CART 算法,它是基于信息熵的决策树算法。信息熵在信息论中称为平均信息量,是对被传送的信息进行度量所采用的一种平均值。用数学概率式定义即一组事件 $X_1, \dots, X_r$,以既定概率 $p(X_1), \dots, p(X_r)$ 出现,其平均值 $E(X)$ 就是信息熵,它的值等于每个事件的(自)信息量 $I(X)$ 的数学期望,公式如下:

$$E(X) = - \sum_{r=1}^r p(X_r) I(X_r) = - \sum_{r=1}^r p(X_r) \log p(X_r) \quad (3)$$

以属性 A 为根的信息增益 $I(A)$ 是

$$I(A) = E(E) - E(A) \quad (4)$$

该算法根据属性集的取值选择实例的类别。它的核心是在决策树中各级节点上选择属性,用信息增益作为属性选择标准,使得在每一非叶节点进行测试时,能获得关于被测试例子最大的类别信息。使用该属性将例子集分成子集后,系统的熵值最小,期望该非叶节点到达各后代叶节点的平均路径最短,使生成的决策树平均深度较小。算法通过计算每个属性的信息增益,并从中挑选出信息增益最大的属性作为给定集合的测试属性,并由此产生相应的分支节点。所产生的节点被标记为相应的属性,并根据这一属性的不同取值分别产生相应的(决策树)分支,每个分支代表一个被划分的样本子集^[11]。

1.2.3 AdaBoost

集成学习是机器学习的首要热门方向。集成学习是使用一系列学习器进行学习,并使用某种规则把各个学习结果进行整合从而获得比单个学习器更好的学习效果的一种机器学习方法。决策树、人工神经网络、朴素贝叶斯分类器等方法是基于单个分类器模型,而集成学习的思路是在对新的实例进行分类时,把若干个单个分类器集成起来,通过对多个分类器的分类结果进行某种组合来决定最终的分类,以取得比单个分类器更好的性能。

著名的 Bagging 和 Boosting 算法都是很好的集成学习方法。目前,Boosting 算法^[12]是集成学习中研究得最深入的一个算法族,而 AdaBoost 算法是 Boosting 家族最具代表性的算法。

AdaBoost 是一种迭代算法,其核心思想是针对同一个训练集训练不同的分类器(弱分类器),然后把把这些弱分类器集合起来,构成一个更强的最终分类器(强分类器)。其算法中不同的训练集是通过调整每个样本对应的权重来实现的。开始时,每个样本对应的权重是相同的,即样本个数为 n ,初始化时对每个训练样本赋相等的权重 $1/n$,在此样本分布下训练出一弱分类器。对于分类错误的样本,加大其对应的权重;而对于分类正确的样本,降低其权重,这样分错的样本就被突出出来,从而得到一个新的样本分布。在新的样本分布下,再次对弱分类器进行训练,得到弱分类器。依此类推,经过 T 次循环,得到 T 个弱分类器,把这 T 个弱分类器按一定的权重叠加(boost)起来,得到最终想要的强分类器。

2 实验结果与讨论

2.1 实验数据与环境

本文共采集 400 例 CLSM 皮肤图像,其中黑色素瘤和良性痣(即正负样本)各为 200 例,所有图像经过病理切片得到确诊。为了方便分析,将所有图像大小归一化为 512×512 像素。图像的小波变换所选取的小波基为 db4,小波分解尺度 $L=5$,因此可得到图像依次从低频到高频的 13 个频率带信息(前 4 个低频率带合为一个低频带)。训练数据样本为图像各个频率带的标准差、能量、熵值:StandardDeviation0 ~ StandardDeviation12, Energy0 ~ Energy12, Entropy0 ~ Entropy12, 共 39 个属性值。实验中采用十倍交叉验证方法,即将所有数据均分成 10 等份,每次取其中 9 份作为训练样本,余下 1 份作为测试样本,重复 10 次实验后计算总的结果进行最后的分析比较。实验采用的计算机为 Intel Core2 Duo E7500 2.93 GHz

CPU,4 GB 内存,32 位操作系统,编程环境为 MATLAB R2010a。

2.2 分类性能比较

本文选取常用的几种分类算法进行实验,首先对数据集构建 CART 分类回归树,设定停止建树的标准是节点上的样本数为 1。当节点数为 7 时取得最小相关损失代价,算法最优分类结果如表 1 所示;其次采用 ID3 决策树构建算法,分类结果如表 2 所示;最后采用 AdaBoost 算法,通过对 CART 进行加强。实验中比较不同参数训练出的对预测数据集的识别率来选择一个较好的参数 σ^2 。表 3 给出了 AdaBoost 方法分类结果 ($\sigma^2=0.5$)。其中:类别 1 代表黑色素瘤,类别 2 代表良性痣;正确分类率,即分类器的正确率指分类器正确分类的项目占有所有被分类项目的比率;敏感性,即病人中得出阳性检测的样本占病人总数的百分比;特异性,即健康人中得出阴性检测的样本占健康人总数的百分比。

表 1 CART 算法分类结果

类别	图像总数	正确分类率/%	类别 1 个数	类别 2 个数
1	200	94.50	189	11
2	200	90	20	180

表 2 ID3 算法分类结果

类别	图像总数	正确分类率/%	类别 1 个数	类别 2 个数
1	200	90.50	181	19
2	200	89	22	178

表 3 AdaBoost 算法分类结果

类别	图像总数	正确分类率/%	类别 1 个数	类别 2 个数
1	200	96.50	193	7
2	200	93	14	186

从表 1、2 可以看出,CART 算法与 ID3 算法的分类正确率相比,特异性相当,CART 算法敏感性偏高。表 3 为 AdaBoost 算法识别结果,两种分类正确率都明显高于前两种算法。

图 2 为 AdaBoost 算法一次实验仿真结果,采用 10 倍交叉验证方法,横坐标为迭代次数,纵轴为测试错误率。可以看出,三种 AdaBoost 算法通过迭代后错误率稳定,均小于 0.07;modest AdaBoost 算法效果最好,迭代 100 次后测试错误率为 0.06。最后对三种机器学习方法分类性能进行比较,正确率取各算法的敏感性与特异性的均值,对照结果如表 4 所示。

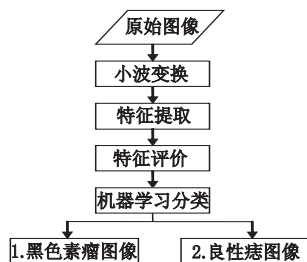


图 1 原理流程

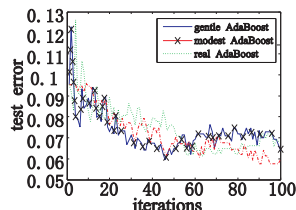


图 2 AdaBoost 算法仿真结果

表 4 CART 算法、ID3 算法、Adaboost 算法分类性能比较

比较项/%	CART 算法	ID3 算法	AdaBoost 算法
敏感性	94.50	90.50	96.50
特异性	90	89	93
正确率	92.25	89.75	94.75

3 结束语

本文把几种常用经典的机器学习方法应用于黑色素瘤的计算机辅助诊断中,并通过实验比较了它们的分类性能。仿真测试结果表明 CART 算法分类正确率比 ID3 算法高,而低于 AdaBoost 算法。通过比较可以看出,AdaBoost 算法在测试中具有较好的分类识别性能,提高了计算机辅助诊断恶性黑色素瘤的准确率,降低了良性黑色素瘤的误诊率。

决策树方法^[13]是分类应用中采用最广泛的模型之一,与神经网络和贝叶斯方法相比,决策树无须花费大量的时间和进行上千次的迭代来训练模型。对于简单的二分类问题,CART 算法具有模型简单、易理解、分类快的优势,但分类正确率不高,当处理类别多时也容易出错。

集成学习^[14]是机器学习中一个热门的研究方向,集成分类器分类效果明显优于单个分类器,但 AdaBoost 算法需要耗费时间复杂度成本,并且需要收集大量的数据集。关于 AdaBoost 算法背后的具体机理,如何提高集成学习效率的问题,以及在这些领域中新的思想、方法,值得在今后的工作中深入研究。

参考文献:

- [1] 李泓馨,张选棉,高天文.“皮肤 CT”——皮肤病诊断的新手段[J]. 中国皮肤性病学杂志,2007,21(7):432-434.
- [2] 魏志平,钟连生,金鑫,等. 皮肤三维 CT 在寻常性银屑病诊断中的应用[J]. 中国皮肤性病学杂志,2011,25(8):607-608.
- [3] 刘春,杨韬,王娟,等. 计算机辅助诊断技术中图像纹理研究的主要方法及其应用[J]. 中国组织工程研究与临床康复,2009,13(39):7721-7727.
- [4] WILTGEN M, GERGER A, SMOLLE J. Tissue counter analysis of benign common nevi and malignant melanoma[J]. International Journal of Medical Informatics, 2003,69(1):17-28.
- [5] GAREAU D, HENNESSY R, WAN E, et al. Automated detection of malignant features in confocal microscopy on superficial spreading melanoma versus nevi[J]. Journal of Biomedical Optics, 2010,15(6):1-10.
- [6] ZHANG Hong, ZHANG Xuan-bing. Texture feature extraction based on wavelet transform[C]//Proc of International Conference on Computer Application and System Modeling. 2010:146-149.
- [7] WILTGEN M, GERGER A, WAGNER C, et al. Evaluation of texture features in spatial domain for automatic discrimination of histologic tissue[J]. Anal Quant Cytol Histol, 2007,29(4):141-149.
- [8] 杨长盛,陶亮. 几种机器学习方法在人脸识别中的性能比较[J]. 计算机工程与应用,2009,45(4):169-172.
- [9] 汪世义,陶亮,王华彬. 几种机器学习方法在 IDS 中的性能比较[J]. 计算机仿真,2010,27(8):92-94.
- [10] 钱捍丽,荀恩东. 基于分类回归树 CART 的汉语韵律短语边界识别[J]. 计算机工程与应用,2008,44(6):169-171.
- [11] 黄文. 决策树的经典算法:ID3 与 C4.5[J]. 四川文理学院学报, 2007,17(5):16-18.
- [12] 左登宇. 基于 AdaBoost 算法的人脸检测研究[D]. 合肥:中国科学技术大学,2009.
- [13] 刘莺迎. 决策树分类算法的分析和比较[J]. 科技情报开发与经济,2008,18(2):65-67.
- [14] 韩彦霞. 基于选择策略的集成学习方法研究[D]. 保定:河北大学,2011.