

机器学习辅助肿瘤诊断研究

杨克戎

(遵义医科大学, 医学信息工程学院, 贵州 遵义 563000)

摘要: 机器学习有多种优势, 其最先用于语音识别及图像辨认等工作中, 能够在短时间内提取关键信息。随着研究的不断深入, 一些学者指出机器学习可以为现代医学效力, 结合人工诊断方式, 能够全面提升诊断结果的真实可靠性。研究发现, 机器学习在肿瘤诊断中的应用成为当前医学发展的重要趋势, 通过图像识别技术与算法模型的应用, 可以在无人工干预的情况下自动对患者进行肿瘤诊断, 极大程度上提升了医疗真诊断效率。本文重点论述人工神经网络与深度学习的概念, 总结机器学习应用情况, 提出其在乳腺癌识别中的应用策略, 仅供参考。

关键词: 机器学习; 肿瘤诊断; 研究

一、人工神经网络与深度学习

(一) 人工神经网络

人工神经网络模拟人体大脑的思维模式, 从一点出发, 能够引申出多个内容。从其构建的知识模型来看, 总体上呈现非线性分布式规律。客观来讲, 人工神经网络可以视为人体大脑中神经元的分级突触组织, 其接收信息的途径比较广泛, 同时还能有效识别各个信息并将其分类。从信息传导模式来看, 人工神经网络能够对收集到的信息进行加权, 并通过计算加权总值判断信息传导效率。从人工神经网络的逻辑单元组成情况来看, 其有三个层级, 分别为入层(主要负责数据信息的收集与整理)、隐藏层(主要负责多层数据的合成)与输出层(数据信息的最后加工层, 能够归纳数据并生成特定的数据报表)。人工神经网络有多种优势, 主要体现在其数据处理功能上, 其能在短时间内快速识别线性类型的数据, 同时也能构建相应的数据模型, 对于混杂及不完全数据的处理, 其具有高容错能力。此外, 人工神经网络具有很强的模拟性, 能够模拟真实的数据模型, 并在此基础上进行整体模型稳定性的预估, 最终判断出潜在的问题, 于数据准确率的提升有积极意义。

(二) 深度学习

深度学习区别于传统意义上的浅层学习, 其曾被评选为2013年度十大科技突破创新技术之一, 同时也可以视为一种深层的人工神经网络。从其数据模型来看, 其具有多个层次, 且数据呈现非线性分布式规律, 整个模型能够自动检测系统中的原始数据, 特别是像素、字符等, 并依据这些内容特征进行分类。与传统人工神经网络相比, 深度学习的层次更多, 特别是隐藏层的数量, 其能在短时间内根据隐藏层存储的数据特征对其进行简单的合并, 避免数据冗余。从深度学习的算法结构来看, 其能自动提取分类需要的低层次或高层次模块中的数据。不仅仅是针对数据处理, 深度学习对于图像资料的处理也有显著优势, 首先, 利用相关算法对原始图像图层进行简单运算, 在此基础上得到一个低层次表达(例如边缘检测器、小波滤波器等), 在低层次表达基础上再重复运算, 通过累加运算结果最终得到一个高层次的表达。

总体来看, 深度学习具有很强的调整功能, 其根据研究问题及数据特征灵活调整层数, 这种情况下能够合并一些重复信息, 同时也能提升整体的学习效果。随着研究的不断深入, 深度学习的分支也不断增多, 如卷积神经网络、循环神经网络等。与传统机器学习语言对比, 深度学习便捷性特征比较明显, 主要体现在其仅需求小部分人工编程, 就可以自动识别信息源, 并对数据进行再分析,

进而进行信息的传递。此外, 深度学习的隐藏层比较多, 能够容纳多组数据, 这也为数据模型的构建提供了多元化的渠道, 能够显著提升整体模型的信息识别与分类能力。从本质上看, 深度学习模式可视一种监督式学习模式, 在机械学习过程中能够根据特定的算法纠错, 且生成特定的数据报表, 最终的结果包含(0, 1)。为了优化学习路径, 深度学习还能推算计算离目标点最近的路径, 进而达成既定的目标。另外, 深度学习具有对混杂因素不敏感, 仅对特定目标特征敏感的优点。

二、机器学习应用现状

机器学习应用的领域比较广泛, 从本质上看, 其是一门多领域交叉的学科, 其依托计算机为媒介, 模拟人类的学习行为。从其学习特征来看, 高效、低耗特征比较显著, 同时也能提升数据的准确性。目前来看, 机械学习在各个领域中均有重要应用, 且得到了社会各界人士的广泛关注。现阶段一些医学领域开始了机械学习的研究, 主要用于一些病理特征图片分析方面, 特别是计算机辅助诊断, 与临床医学建立了有机联系, 也能提升诊断结果的真实可靠性。

三、支持向量机(SVM)算法及其在乳腺癌识别中的应用

(一) SVM算法

SVM是一种经典的机器学习算法, 其基础是Vapnik根据统计学习理论提出的VC维和结构风险最小化准则, 其能最大限度的减少误差, 在此基础上降低安全风险。从数据处理特征来看, 其容错率比较高, 同时对于数据集也没有维数限制。在进行线性分类时, 要适当调整分类超平面设置的距离, 使其位置与两类样本距离比较远。在进行线性近似可分时, 要转变以往的分析理念, 可引入松弛变量理念, 能够更好地解决少数异常样本不可分的情况。在进行非线性分类过程中, 为了提高数据处理效率, 相关人员可借映射函数之力, 实现数据的维度转变, 由低维转化至高维, 使非线性分类变成高维空间的线性分类模式。

(二) 理论基础

1. VC维理论

VC维理论与数学思想高度契合, 其最先始于数学中的统计学习理论, 运用数学语言阐述变量之间的关系, 使学习者通过数学分析了解系统表达的内容。目前来看, VC维理论已经得到国内外学术界的一致认可, 同时也为诸多机器学习模型的性能解释提供了理论支撑。总体来看, VC维理论具有很强的数据分析与处理能力, 能够整合原有数据, 并将其按照一定的算法归纳整理, 生成相应的数据模型, 进而更好地预测数据分布规律。随着时代的发展,

VC 维被应用于多个领域，其主要功能为进行性能评判。在医学领域中应用该理念能够更好地评价医疗设备性能，也能从一定程度上提升诊断结果的准确性。

2. 结构风险最小化准则

结构风险最小化准则是由 Vapnik 提出的，其能在短时间内实现多组数据的快速整合，便于专业人员重新调整结构，这种情况下也能有效降低结构风险。按照函数的计算法则，将函数集 S 分为多个函数子集，如下图 1 所示：

$$S_1 \subset S_2 \subset \dots \subset S_n \dots$$

图 1

从上图可以看出，分解后的子集既可以根据置信范围值进行排序，这种情况下也能了解各个元素对应的 VC 维。在此基础上，根据子集种类进行复杂度判断，进而判断整个系统的结构风险（子集复杂程度越高，系统安全性越高）。最后，相关人员要进行计算经验风险范围的计算，可按照图 2 方式进行相关计算，分析子集特征，在此基础上选择风险最小的函数作为最优函数。

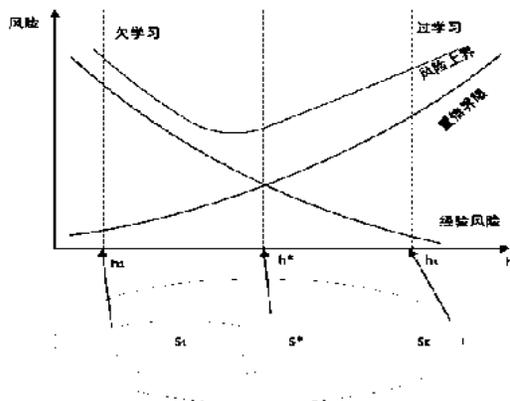


图 2 结构风险最小化准则示意图

从上图所示的算法模型中，我们可以直观看出 VC 维、经验风险、置信范围和风险值边界的关系。VC 维增加的情况下，整个系统的安全性也会提升，这种情况下也能提升整个系统的信息识别能力。如 VC 维数值增加 h 时，置信范围也增大，且超过一定数值时危险性也容易增大。相反，如减少 VC 维数值，整个系统的信息识别能力也大大降低，这种情况下数据的真实性不能保证。如样本数据处于 n 且固定不变时，经验风险和置信范围是一组矛盾体，没有同时降低的区间，但也存在使整体结构风险最低的 VC 维值。

(三) SVM 算法识别乳腺癌的流程及缺点

1. SVM 算法识别乳腺癌的流程

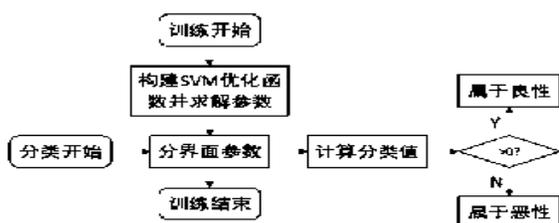


图 3 SVM 算法识别乳腺癌流程图

2. SVM 算法缺点

SVM 算法为最具代表性的机器学习算法，其研究历程比较久，且在多个领域中均有重要应用。上个世纪九十年代中期，SVM 算

法被评为最有价值性的算法，且为各个领域数据分析及设备性能判断提供了有价值的依据。直到现在，SVM 算法也被应用于各大领域，发挥了重要作用，其性能比较优越，但在实践应用过程中也存在一些问题，主要体现在以下三个方面：

(1) 不能快速处理大规模数据。虽然 SVM 在小规模数据处理中发挥了重要作用，能够比较好地梳理样本数据之间的关系，也能得出相应的模型，以便研究人员更好地探究不同数据之间的线性或非线性关系。但在数据比较多的情况下，该算法的弊端也逐渐显现出来，其不能在短时间内实现海量数据的处理，这是由于其受固定算法的约束，在大规模数据处理中显得无力。

(2) 不能直接进行数据分类识别。从数据分离识别模式来看，SVM 算法多采用间接方案，如不使用这种方案不能实现多重数据的快速整合。

(3) 存在制约识别精度的超参数。SVM 算法中存在制约识别精度的超参数，即惩罚因子 C 和核函数参数 g。这两个参数无法使用概率法求得，只能通过穷举策略或者优化算法确定。

四、结语

机器学习有诸多优势，作为人工智能发展的重要板块，在现代医学中已经展现出重要的发展潜力，尤其在辅助肿瘤诊断方面中发挥了重要优势。自 20 世纪 90 年代起，医学领域就开始应用图像识别技术进行诊疗研究，例如，美国针对皮肤癌开发了具有人工智能属性的检测系统，在临床验证中发现，其检测结果与专家诊断结构具有较高的重合率，同时还展现了其更高的特异性与灵敏度。由此可见，机器学习具有高效率、高准确性等诸多优点，其在医学领域中的应用拓宽了医学研究内容，在后续研究过程中发现其在肿瘤基因检测中也有重要应用，其通过图像识别技术以及各类算法，构建了相应的数据模型，可以自动识别肿瘤，从一定程度上提升了诊断效率。本文以乳腺癌为典型，探讨机器学习在该疾病诊断与治疗中的应用，可见机器学习的优势。在应用过程中，还出现一系列问题，需要更多研究人员深入相关研究中，优化诊断档案，使现代科技为医学效力，进而带动整个医学领域的发展。

参考文献：

[1] 梅宏翔, 程俊豪, 李一洲, 马凤税, 张恺文, 寿宇珂, 李杨. 机器学习在颌面部囊肿及肿瘤中应用的研究进展 [J]. 华西口腔医学杂志, 2020, 38 (06): 687-691.

[2] 董竞方. 机器学习在甲状腺恶性结节超声图像识别及中医舌象分类中的应用 [D]. 北京中医药大学, 2020.

[3] 潘德润. 基于机器学习模型探讨 X 线征象在骨肿瘤分类中的价值及其临床应用 [D]. 南方医科大学, 2020.

[4] 张佳佳, 樊鑫, 秦珊珊, 余飞. 基于深度学习的人工智能在肿瘤诊断中的应用进展 [J]. 国际放射医学核医学杂志, 2020 (01): 11-15.

[5] 李珂. 机器学习算法在眼角膜与骨肿瘤影像中的分析评估研究 [D]. 北京邮电大学, 2020.

[6] 许凯波, 罗广祥, 孙震辉. 基于深度信念网络的乳腺肿瘤辅助诊断研究 [J]. 计算机与数字工程, 2019, 47 (03): 582-586.

[7] 赵可扬, 杨沐月, 朱静好, 王泽淇, 沈炜炜. 机器学习辅助肿瘤诊断 [J]. 肿瘤, 2018, 38 (10): 987-991.