

核医学中的机器学习:第一部分——引言

陈文坤(译) 李雪娜 李亚明(审校)

中国医科大学附属第一医院核医学科,沈阳 110001

本文首次发表在 *The Journal of Nuclear Medicine*, 2019, 60(4): 451-458

Machine learning in nuclear medicine: part 1—introduction

Carlos F. Uribe¹, Sulantha Mathotaarachchi², Vincent Gaudet³, Kenneth C. Smith⁴, Pedro Rosa-Neto², François Bénard^{1,5}, Sandra E. Black⁶, Katherine Zukotynski^{7,8}

¹Department of Molecular Oncology, BC Cancer, Vancouver, British Columbia, Canada; ²Translational Neuroimaging Lab, McGill University, Montreal, Quebec, Canada; ³Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada; ⁴Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada; ⁵Department of Radiology, University of British Columbia, Vancouver, British Columbia, Canada; ⁶Department of Medicine (Neurology), Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada; ⁷Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada; ⁸Departments of Medicine and Radiology, McMaster University, Hamilton, Ontario, Canada

Corresponding author: Katherine Zukotynski, Email: zukotykn@mcmaster.ca

原文 DOI:10.2967/jnumed.118.223495

【摘要】 该系列文章介绍了核医学背景下的机器学习(ML)。第一部分讨论了 ML 的历史,描述了常见的算法并举例说明 ML 何时可用于核医学。第二部分聚焦于 ML 在当下对核医学领域的贡献,讨论了未来对 ML 的预期和局限性,并对 ML 能做什么和不能做什么进行非常严谨的评估。

【关键词】 机器学习;人工智能;核医学;算法

DOI:10.3760/cma.j.cn321828-20210624-00205

目前,几乎每次会议都有关于人工智能(artificial intelligence, AI)和机器学习(machine learning, ML)的内容,相关术语也越来越多地出现在文献中。AI 一词最早出现于 1956 年,指那些执行往往需要人类智能(如视觉感知或决策)参与任务的计算机。ML 是 AI 的一个分支,可追溯到 20 世纪 50 年代末的电子计算^[1],其在 20 世纪 60 年代应用于医学^[2]。

广义上讲,ML 指的是一种算法,通过编程,从观察中学习,然后根据所学内容而不是程序员的感性知识进行统计推断。传统上,程序员特定了所有算法,除了已编写的内容之外,没有什么可以进一步学习的。但 ML 算法在编程时包含了具有未知值的参数,这些参数是通过向程序反馈观察结果来学习的。

为什么 ML 算法会在当下的医学成像领域中兴起?原因包括理论的进步(更好的算法)^[3]、微电子学(更好的硬件)^[4]以及大量可用的训练数据(即大数据)。此外,放射组学和放射基因组学等领域也显示出了巨大的潜力^[5]。放射组学是指提取和量化图像纹理、形状和其他质量的数学描述符(称为特征),这些特征可以与多种数据相关,如组织病理学发现和临床结果(生存率)。放射基因组学将放射组学特征与基因组数据更明确地联系起来。在这些新领域使用 ML,通过识别以前未知的关系来研究科学现象。

本文是关于核医学中 ML 系列文章的第一部分,目的是介绍 ML,讨论常见的算法并强调它们之间的差异。

一、ML 算法的鸟瞰图

ML 算法包括的核心元素有数学模型、成本函数和数据。

1. 数学模型。数学模型是解释数据之间的关系并将输入的观测值转换为输出值的一种方法。例如,数学模型可以是一个函数,它能将来自图像的输入特征[如脑淀粉样蛋白 PET 显像在感兴趣区(region of interest, ROI)中的标准摄取值比值(standardized uptake value ratio, SUVR)]转换为输出决策(如淀粉样沉积的扫描分类为阳性或阴性)。通常基于准确性、任务适用性和实际实施三者间的权衡来决定使用哪种模型。模型输出观测值为连续值(如淀粉样沉积的量)的问题称为回归问题;输出观测值为离散值或类别(如淀粉样蛋白阳性或阴性)的问题称为分类问题。不同的数学模型具有不同的复杂程度或预测能力。许多模型都有超参数,这些参数在训练之前选择,但如果性能不足则可以调整。例如,最初可能规定任何给定的脑淀粉样蛋白 PET 显像结果被分类为淀粉样沉积阳性或阴性(允许 2 个结果);然而训练可能表明某些扫描结果是模棱两可的(允许 3 个结果)。

2. 成本函数。成本函数(或损失函数)用来衡量模型性能与其预期结果的接近程度。例如,成本函数可以是误分类率或均方误差。

3. 数据。正如医师通过回顾病例来学习一样,ML 算法使用数据进行训练。数据可以是图像测量(如淀粉样蛋白 PET 显像研究的 SUVR)、存活数据或这些数据与其他观察的组合。数据通常分为训练、验证和测试子集。训练数据用于模型教学,验证数据用于超参数微调,测试数据用于确认训练模型的性能。通常,数据包括已知的输出,如医师可能已经将所有可获得的淀粉样蛋白 PET 显像分为淀粉样蛋白阳

性或阴性;但有时输出是未知的,如可能已知所有 ROI 的 SUVr,但不知道显像是淀粉样蛋白阳性还是阴性。根据数据是否包含已知输出将 ML 算法分为两大类:有监督学习和无监督学习。

二、有监督学习和无监督学习

有监督学习是指使用已知输出训练的 ML 算法。例如, ML 算法可以利用先前的 SUVr/ROI 结合的和神经病学发现来确定神经系统的结果。从某种意义上说,有监督的算法根据训练数据确定数学模型的未知参数,以预测未来测试数据的结果。下面描述的有监督的 ML 算法的示例有线性回归、支持向量机、随机森林和人工神经网络 (artificial neural networks, ANNs)。

无监督学习是指使用未知输出的训练数据的 ML 算法。这些算法通常解决诸如识别数据点之间的共性(如在无 PET 显像所属类别的先验知识的情况下将淀粉样蛋白 PET 显像聚类成淀粉样蛋白阳性或阴性)或寻找压缩数据的方法(如减少预测结果所需特征的数量)等任务。下面描述的无监督 ML 算法的示例有 k 均值聚类、主成分分析和奇异值分解。

图 1 从理论上说明了这些概念。在图 1A 和 1B 中,数据点分为圆形和三角形;在图 1C 中,数据点未分类。如果目的是训练 ML 算法,将数据适当地分类为圆形和三角形,并且分类方案已知,则首选有监督的方法。对于分类方案未知的场景,将数据分门别类则需要无监督的方法。一个观察者可能会说这些数据点有 2 个聚类簇(左下和右上)。然而,另一个观察者可能会将数据点分组到更多的簇中。归根结底,对类别的数量进行初步的有根据的猜测可能是必要的。图 1D~1F 显示了应用于每个场景的潜在模型。在图 1D 中,直线在圆形和三角形之间创造了一个分界。这种情况下的数学模型是线性的,即 $y=ax+b$, ML 算法必须学习 a 和 b 的值,以使误分类的数量最小化(如使成本函数最小化)。在图 1D 中,直线 2 的成本低于直线 1,因为圆形区域中的三角形数量较少。一个有效的 ML 算法可以使 a 和 b 适应更好的解。在图 1E 中,线性模型将不可避免地导致大量的错误分类。在这种情况下,更复杂的非线性模型可能更适合此任务。在图 1F 中,一个算法将数据分类为 2 个聚类簇。下面讨论了几种常见

的 ML 算法,并在表 1 中进行了总结。

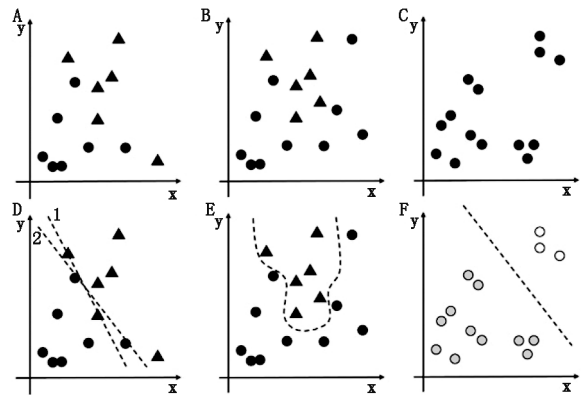


图 1 机器学习任务。A,B.输入测量被分为圆形或三角形用于有监督学习(如三角形可以是阳性病例,圆形是阴性病例);C.未分类的输入测量(如临床信息未知);D.直线 2 比直线 1 更适合用于分类;E.需要非线性方法分类;F.潜在聚类簇

三、常见的 ML 算法

ML 算法有许多(图 2~7),通常用 R、Matlab 和 Python 等语言编写^[6],其中包括内嵌支持(如库、工具箱或包)。这些算法在许多方面有所不同,并且最佳算法的使用通常取决于任务、数据和数学模型的复杂性(如要执行的加法和乘法等操作的数量)。

1.朴素贝叶斯分类。朴素贝叶斯分类是一种有监督算法,它对遵循简单概率分布(如高斯分布或多项式分布)的特征进行分类,并假设这些特征彼此独立(允许数学简化)。超参数是分布参数(如每个类别特征值的均值和方差)。通过查看特征并计算能使概率函数最大化的类来完成分类,其计算成本低。Mehta 等^[7]的一项回顾性研究使用了多项式朴素贝叶斯分类,从¹⁸F-脱氧葡萄糖 (fluorodeoxyglucose, FDG) PET/CT 的特征预测了对⁹⁰Y 放射性栓塞治疗的反应,使用 22 个训练集和 8 个测试集,预测反应的灵敏度约为 80%。

2.线性回归。线性回归是一种用于回归问题、有监督的算法,其数据集包含 n 维连续特征,并且模型是最适合数据集的 n 维超平面(因此是线性的)。成本函数定义为真实结

表 1 机器学习算法的关键方面

算法	任务	是否监督	模型	典型成本函数	计算负担	假设/评论
朴素贝叶斯分类	分类	有监督	几个(如高斯分布)	概率	低	依赖于朴素概率分布
线性回归	回归	有监督	超平面	均方误差	低	
支持向量机	分类或回归	有监督	超平面	分类率,均方误差	中等	处理复杂问题
随机森林	分类或回归	有监督	树	分类率,均方误差	低~中等	容忍过度拟合
人工神经网络	分类或回归	有监督(典型);无监督/强化学习(不太常见)	神经元层层连接	分类率,均方误差	高	用于复杂的问题;可能是卷积的或深层的
k 均值聚类	聚类	无监督	簇质心	到簇中心的距离	中等(取决于问题)	识别质心并将数据分配给最近的质心
分层聚类	聚类	无监督	树状图	数据点之间的距离	中等(取决于问题)	通过识别相似的数据点对数据进行聚类
主成分分析	降维	无监督	主要成分		中等(取决于问题)	

果与模型预测结果之间的均方误差,计算成本低。

3. 支持向量机(图 2)。支持向量机是一种有监督的算法,用于二元分类或者回归,后者不太常见。该算法识别出一条曲线(如果有很多维度则为超曲面)通过使每个类中最接近的点之间的距离最大来更好地区分 2 个类。离直线最近的数据点称为支持向量,算法因此得名。一旦经过训练,对于较小的数据集来说,支持向量机计算效率高(如一旦经过训练,算法只需要确定新的数据点位于曲线的哪一侧)。支持向量机还可以通过移动到更高维空间、使用非线性数据转换或使误分类的数量最小化来处理类之间无法绘制清晰的线性分离(超平面)数据(图 1B)。如 Van Weehaeghe 等^[8]基于¹⁸F-FDG PET 显像中脑代谢的差异,使用支持向量机将受试者分为肌萎缩侧索硬化(amyotrophic lateral sclerosis, ALS)患者和健康对照组。所有 PET 数据在空间上被标准化,并使用感兴趣体积和基于体素的统计参数映射协议分析进行定量分析。一个由 70 例 ALS 患者和 20 名健康志愿者组成的队列研究使用基于体素的比较和脑组织分割掩码来训练支持向量机(线性模型),基于超过¹⁸F-FDG PET 图像平均值 50%的体素定义脑组织分割掩码,随后 105 例 ALS 患者作为测试数据,结果显示支持向量机对分类的准确率为 100%。由于 ALS 患者在训练和测试数据中的脑代谢模式相似,支持向量机能够正确区分 ALS 患者和健康对照者。几乎不需要探索一种计算强度更大的 ML 算法来解决这个任务。该研究还招募了 10 例原发性侧索硬化患者。由于 ALS 患者和原发性侧索硬化患者的脑代谢相似,因此无法对这些患者进行鉴别,需要进一步的评估来对这些患者进行分类。

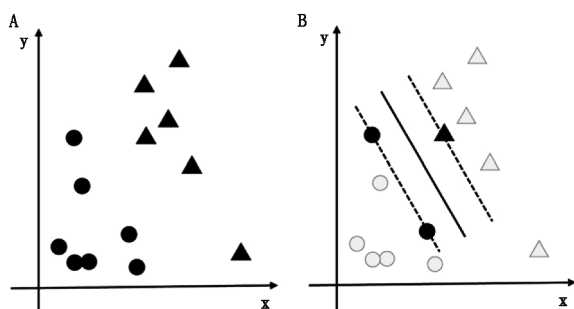


图 2 支持向量机的图解。A. 训练数据集显示为散点图; B. 实线表示每一类中最近数据点的最大距离, 实心点表示最近的点, 称为支持向量; 虚线之间的区域中没有数据点

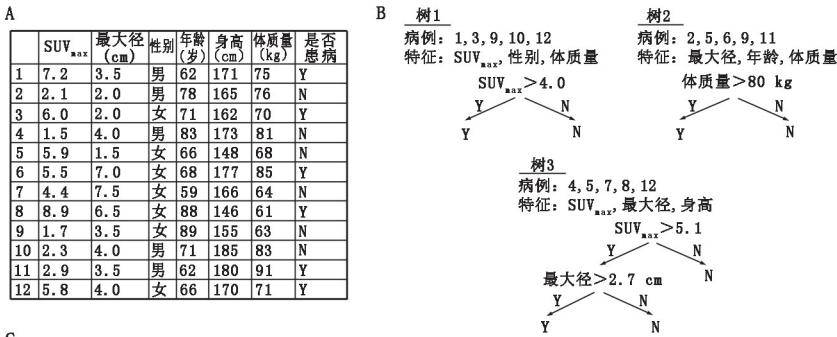
4. 随机森林(图 3)。随机森林是一种有监督的算法, 其创建了一个决策树集合(即森林)用于分类或回归。例如, 人们可能希望通过学习一系列已知神经病学发现的患者的淀粉样蛋白 PET 显像的 SUVR/ROIs 来预测神经病学发现。为了建立森林图中的每个决策树, 算法选择一个随机的病例子集和一个随机的特征子集(这些子集的大小是超参数)。如病例子集是从临床读取的淀粉样蛋白 PET 显像中随机选择的, 特征子集是从可用的 ROI 中随机选择的。每个决策树都是通过识别特征(如 ROI)和特征值(如 SUVR)从根节点开始递归地构建, 根据误分类率能更好地将子集的病例分成左子树和右子树; 然后, 以相似的方式创建 2 个分支(左分支对应左子树, 右分支对应右子树)。树的端点表示输出, 当子树

的所有病例都来自同一类(如全部为阴性或全部为阳性)时就到达了树的端点; 端点可以是类(对于分类问题), 也可以是数值(对于回归问题)。每个决策树都是不同的, 因为其处理不同的数据子集。一旦森林被创建, 新的数据就会呈现给每个决策树进行分类; 最终的结果是所有树的决策中最常见的那个决策。如最终结果可能是最常见的临床发现(类); 淀粉样沉积阳性或阴性。最近的一项研究表明, 在一组 366 例肝内肿瘤患者中, 可以训练一个有 2 000 个决策树的随机森林, 根据年龄、肝功能和肿瘤负担等因素预测⁹⁰Y 放射性栓塞后的患者死亡率^[9]。该研究使用了最小深度(即在变量中从树的主干到第一次分支的最短距离)来识别预测性测量(基线胆碱酯酶和胆红素), 一旦胆红素水平超过 15 mg/L, 术后死亡率就会显著增加。

5. ANNs(图 4)。ANNs 通常(但并不总是)受监督, 可用于分类或回归问题。ANNs 逐步地处理数据(称为层)。在每一层, 几个简单的计算单元称为神经元(受生物学启发), 它们处理输入数据并将结果传递给下一层的神经元(图 4A)。神经元的计算通常包括输入的加权求和以及偏倚, 然后是结果的非线性转换(如阈值化)。诸如反向传播之类的算法用于训练(如学习加权值)^[10]。在 ML 算法中, ANNs 往往是最强大的, 但也是最具计算挑战性的。通常, 一层中的每个神经元都连接到下一层中的每个神经元, 设想一个可处理 100×100 像素图像的 ANNs, 例如有 10 000 个输入。如果每一层的神经元数目与输入数目一样多, 则每一层需要 10 000×10 000=10⁸ 个连接, 每个连接至少需要 1 个乘法和加法运算, 因此每一层的计算成本很高。在 ML 的早期, ANNs 受到了高计算成本(当时的微电子技术无法解决)的限制并且事实上一些问题仍未解决。最新的发展使 ANNs 能够克服这些障碍, 并且 ANNs 在当下非常流行。

深度学习于 1986 年首次被提出^[11], 2006 年被应用到 ANNs 中来形成深度神经网络(deep neural networks, DNNs)^[12]。DNNs 本质上是具有许多层(10~150, 取决于任务)的 ANNs。由于有大量的层, DNNs 的计算量更大, 但能够处理复杂的问题。此外, 每个 DNNs 层被认为是在不同的抽象级别上处理数据, 如 DNNs 可以将第一层用于图像分割, 然后将下一层用于分割的 ROI 内的病变识别。

借助每个神经元连接到下一层神经元的子集这一结构可以降低每层的高计算成本。图像处理应用的一种方法是卷积神经网络(convolutional neural network, CNN)。CNN 是 DNNs 的一种形式, 其主要区别在于卷积层的使用。具体地说, 在 CNN 中, 一层神经元的输出被认为是图像; 对图像进行二维卷积, 然后进行非线性操作(称为修正线性单元)和形成像素数据池(将几个像素合并为一个像素, 这也降低了成本)。假设在上面的例子中使用大小为 10×10 的卷积掩模, 则只需要 10 000×100=10⁶ 个乘法和加法, 这是 100 倍的节省。图 4B 描绘了一个 2 层的 CNN, 其中每一层执行卷积、修正线性单元和池化。在这些层之后, 另一层神经元(每个神经元都与前一层的每个神经元相连接)计算输出分类。此外, CNN 不仅可用于直接处理图像, 也可用于预定义的 ROI 或感兴趣体积。因此, 没有必要决定提取哪些放射组学特征, 因为 CNN 可以自己选择特征。2012 年, 一个名为 AlexNet^[13]



C 新病例: SUV_{max}=5.3, 最大径=3.5 cm, 男性, 年龄=62岁, 身高=171 cm, 体质量=72 kg
树1: Y, 树2: N, 树3: Y, 最终是Y

图 3 随机森林的图解。A. 训练数据集, 包括 12 个病例 (每个病例有 6 个特征) 和其监督训练的结果; B. 3 个决策树, 每个决策树由 5 个随机选择的病例和 3 个随机选择的特征生成, 对于树 1 和树 2, 1 个特征足以完全分离 5 个病例, 对于树 3 则需要 2 个特征; C. 新病例在树 1 和树 3 得到 Y, 树 2 得到 N, 因此, 多数决定是 Y。SUV_{max} 为最大标准摄取值; Y=患病, N=不患病

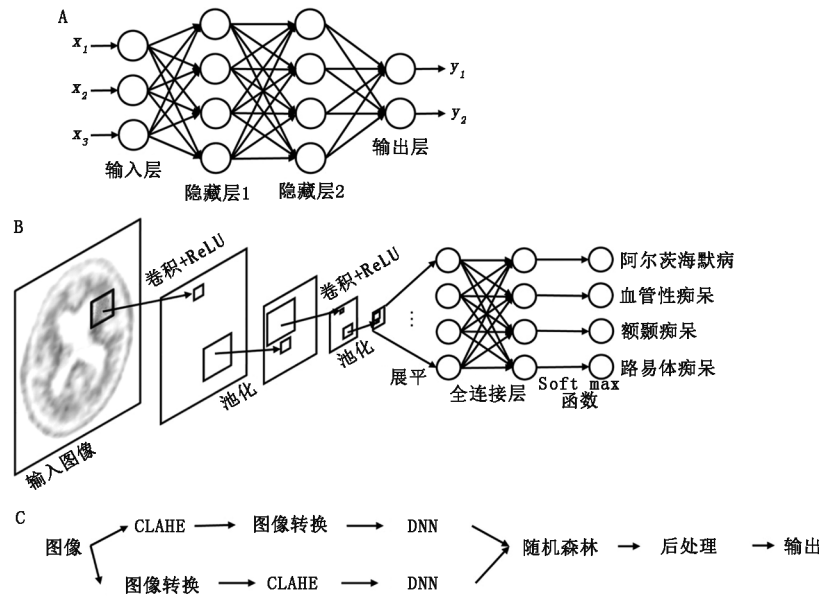


图 4 A. 4 层人工神经网络和神经元 (圆圈), 箭头表示数值从一层传递到另一层。B. 示例的卷积神经网络用于脑部 PET 显像处理。两层分别执行二维卷积, 然后是修正线性单元 (ReLU) 和池化 (取每 4 个像素的最大值以减少像素总数)。生成的矩阵 (特征映射) 被展平成与另一层完全相连的一维神经元阵列。执行非线性操作 (归一化指数函数) 生成分类。为了简单起见, 仅显示单输入分片; 然而, 卷积神经网络可以处理高维或多模式图像。被称为超参数的参数包括层数、卷积滤波器中的像素数、卷积步长和池的像素数。C. Walker 等^[23] 使用的图像处理, 有 2 个能处理不同尺度图像的路径, 每一条路径都具有对比受限的自适应直方图均衡 (CLAHE)、深度神经网络 (DNN) 和之后的随机森林

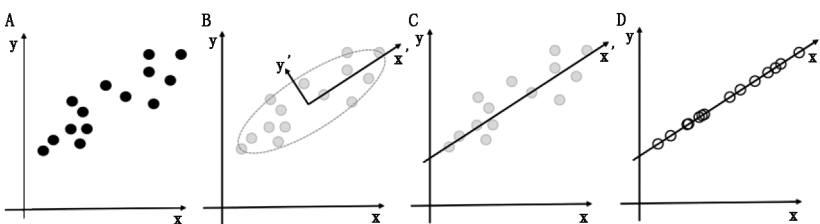


图 5 主成分分析图解。A. 训练数据集显示为散点图; B. 沿旋转轴 x' 扩展的数据大于 y' ; C. 由于组分 y' 的重要性小于 x' , 因此将其去除; D. 数据点沿 x' 轴移动到最近点, 被压缩 (有损) 的数据沿 x' 的测量来表示, 而不是使用 x 和 y 的测量

的 CNN 的推出带来了突破, 其在名为 ImageNet 的开放挑战中超过了先前的 ML 算法。

DNNs 和 CNN (事实上, 在文献中两者几乎没有什么区别, 现今大多数 ANNs 既是深度又是卷积的, 这 3 个缩略词几乎可以互换使用) 共同为 ML 当前的繁荣铺平了道路。一般来说, 与传统的 ML 算法相比, 深度学习训练所需的人工输入更少, 但深度学习也更加复杂, 往往需要大量的训练数据才能有效。在 ImageNet 挑战^[13] 中, 使用了 100 多万带注释的图像。在医学成像中, 往往很难获得如此大量的训练样本。为了绕过这一限制, 研究人员尝试通过使用数据增强技术^[14-16]、降低图像维数^[16-19] 和微调现有的经过预训练的 DNNs^[19-21] 来降低对训练数据的需求。如在某些情况下, 可以通过稍微改变可用数据来增加数据, 比如旋转图像、翻转图像或添加噪声。此外, 如果网络层能够在同一组图像特征与所研究对象相似的图像上进行训练, 那么该学习就可能被迁移到新任务中 (称为迁移学习), 从而减少学习新任务所需的数据。

CNN 已经被应用于多种医学成像设备。如 CNN 与图像预处理和随机森林一起被用于自动系统, 来检测乳房 X 线影像上的恶性肿瘤^[22]。通过使用遗传算法来优化一系列的图像处理步骤 (图 4C)。使用一种称为对比度受限的自适应直方图均衡算法对乳房 X 线影像进行预处理, 接着用 2 个深度的 CNN (每个至少有 20 层), 以 2 种不同的尺度处理图像, 然后用一个随机森林做最后的预测。据报道, 一个包含 7 000 多张图像的数据集被分为训练集、验证集和测试集, 从而生成一个可与放射科医师相媲美的高灵敏度、高特异性的 ML 算法。

6. 主成分分析 (图 5)。主成分分析是一种无监督的数据简化算法。其思想是找到可用于重建初始数据的构建块 (如特征) 的线性组合, 构建块的组合比初始数据简单且能使初始数据的丢失最小化。主成分分析常用于数据压缩。例如在核医学中, 主成分分析可以用于降低 PET 显像中的呼吸伪影^[23], 在这种情况下, 任务需要识别最能表示呼吸的数据, 并在重建过程中使用这些信息来降低呼吸伪影。主成分分析与

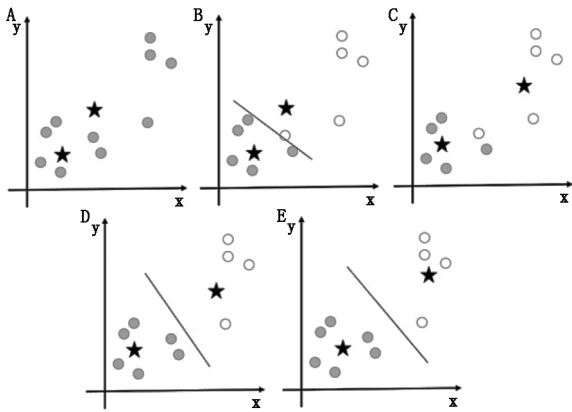


图 6 K 均值聚类的图解。A. 训练数据集显示为散点图(圆圈), 星星表示对于 2 个簇中心的初始猜测; B. 星星中间的实线将数据分成 2 个簇(灰色和白色); C. 星星被移动到每个簇的中心; D. 更新了星星之间的实线; 1 个数据点改变了簇; E. 星星被移动到更新后的中心, 由于没有数据点改变簇, 算法收敛

使用昂贵的外部门控设备获得的结果之间有很高的相关性。线性判别分析与主成分分析密切相关。基本的区别在于, 线性判别分析是一种有监督的算法, 需要构建块的线性组合。

7. K 均值聚类(图 6)。K 均值聚类是一种无监督的数据聚类算法, 其中的聚类是指数据点的集合, 这些数据点在测量值或特征值上彼此接近。K 是由程序员规定的聚类簇的数量。目标是创建 k 个数据聚类簇, 其中每个数据点都被分配给最相似的簇。计算机通过迭代来实现这些: 第一步, 选择 k 个簇中心(即为每个特征选择随机值); 第二步, 将每个数据点分配给中心最近的簇; 第三步, 将簇中心调整成位于那些在步骤 2 中分配给它们的数据点的中心; 第四步, 重复这个过程, 从步骤 2 开始, 在没有数据点能改变簇时结束。考虑到 k 个簇中心最初选择的随机性, 每次运行程序时都有可能产生不同的解。为了说明这一点, 每个已知 SUVR/ROI 组合能够将脑淀粉样蛋白 PET 显像分成 $k=2$ 个簇(如淀粉样蛋白阳性和阴性)。反过来说, 这可能表明哪些 SUVR/

ROI 组合最有助于对显像进行聚类。Blanc-Durand 等^[24]对 37 例用 $O-(2-^{18}\text{F}\text{-}\text{氟}\text{代}\text{乙}\text{基})\text{-L}\text{-}\text{酪}\text{氨}\text{酸}\text{ PET}$ 显像新诊断的胶质瘤患者进行 k 均值聚类分析, 以提示与疾病进程和生存相关的影像学特征。

8. 分层聚类(图 7)。分层聚类是一种无监督的聚类算法。分层聚类通过将距离最近的相邻数据点分为对和聚类簇, 直至没有剩余的自由数据点, 从而来寻找数据的相似性, 由此产生的结构称为树状图。如 Tsujikawa 等^[25]用树状图提示宫颈癌患者¹⁸F-FDG PET 显像的相关特征。

四、需要考虑的问题

其中一个具有挑战性的问题是选择哪种 ML 算法, 或者为什么某个算法能被公开使用。首先, 每个 ML 算法都针对一个特定的任务(如分类、回归、降维或聚类)。如果想使用现有的图像数据库和良/恶性分类来训练 ML 算法对未来的图像进行分类, 那么可以使用有监督的算法, 如朴素贝叶斯分类、随机森林、支持向量机或 CNN。因为用户不太可能知道输入数据的潜在概率分布, 所以可能很快就排除了朴素分类器。接下来必须考虑 ML 算法的复杂性、计算能力和预测能力。例如, 在桌面工作站上执行工作是否可取还是需要更多的计算能力? 事实上, 为了向前发展, 用户可以尝试每一个选项, 看看哪个最有效(使用可用的预编程工具箱), 或者接触那些在 ML 领域有更多理论基础的人来了解下一步工作(如统计学家、数学家、计算机科学家或工程师)。

在选择复杂的 ML 算法之前, 除了潜在的高计算成本外, 另一个还需要考虑的重要因素是过度拟合, 在这种情况下, 数学模型过于紧密地反映了训练数据的微小变化, 而这些变化无法预测验证数据和测试数据(如由于样本数量太少而缺乏概括性)。例如, 考虑图 5A 中的数据, 如果把这看作一个 y 作为因变量的回归问题, 而不是无监督学习, 并且如果为了精确匹配试图用 N 阶多项式来表示数据, 那么就不一定会发现潜在(更简单、更具代表性)的线性趋势。一个好的模型应该具有预测性, 即便其不一定能准确地再现训练数据。虽然详细分析过度拟合的迹象和减少过度拟合的策略

超出了本文的范畴, 但是了解一些对过度拟合具有更高容忍度的 ML 算法(如随机森林)是有用的。

充分训练 ML 算法所需的数据量, 以及获取足够数据的成本高。如上所述, 确实有一些 ML 算法需要非常大的数据集(特别是将其划分为训练、验证和测试子集时); 然而已经出现了许多技术来解决这个问题, 包括数据扩充。数据集的潜在偏倚是一个陷阱。在 ML 早期, 有一些 ANNs 无意中看似无关问题所迷惑, 诸如照明条件。重要的是, 在子群太小时, 即使是最好的增强技术和最宽容的 ML 算法也会失败(这在当前的放射医学和核医学中很常见)。

五、前景展望

尽管 ML 早就存在, 但在医学成像应用还处于初级阶段^[26]。融合成像

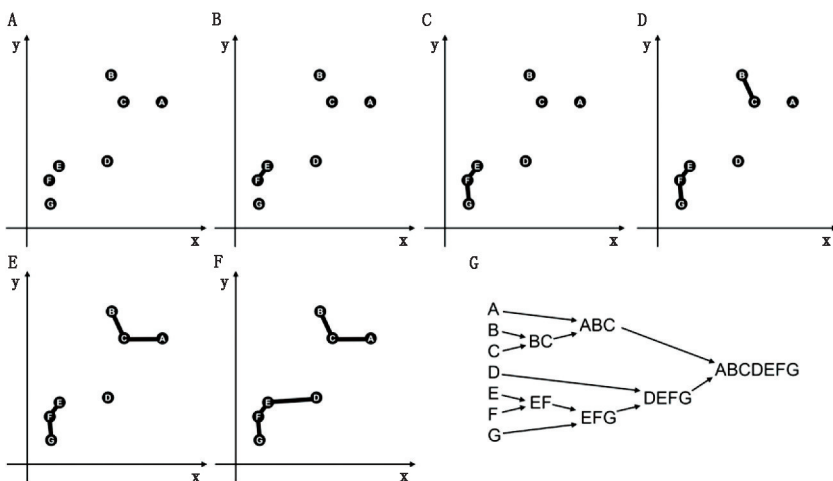


图 7 分层聚类的图解。A. 训练数据集显示为散点图; B. 数据点 E 和 F 相似, 被分配给新的簇 EF; C. 最邻近的是 G 和 EF(G 最接近 EF 中的 F 点), 并被归为 EFG; D. 数据点 B 和 C 相似, 被分配到新的簇 BC; E. 最邻近的为 A 和 BC(A 最接近 BC 中的 C 点), 并归为 ABC; F. 最邻近的是 D 和 EFG, 并被归为 DEFG; G. 树状图表明簇是如何形成谱系的

(如 PET/MR) 的出现给这一领域注入了活力, 因为这种应用非常适合于 ML (鉴于可以同时获得不同类型的信息)。然而, 仍存在一些局限性。通常, ML 算法需要有良好质量的大型医学数据库来提供可靠的结果。2018 年 3 月, Thrall 等^[27] 提议 AI 医学成像应用可从以下 4 个方面受益: (1) 国家和国际图像共享网络; (2) 经证实病例的参考数据集可对 AI 程序进行测试和比较; (3) 用于 AI 应用程序的成像协议的标准化和优化标准; (4) 描述和报告 AI 应用程序的通用词汇。成像数据采集的标准化是非常重要的, 因为用于采集成像数据的扫描规程会影响影像组学特征。此外, 尽管 ML 的初始结果令人兴奋, 但仍需要标准化和方法透明化来提供可重复的结果并加快临床转化^[28]。随着核医学中 ML 的普遍存在, 有几个问题也需要加以明确, 如在患者诊疗中的个人责任、信托契约、数据机密性、无偏倚算法、ML 算法的验证过程以及当结果超出数据所存在的特定患者群体时是否还具有普遍性。

六、结论

本文是系列文章的第一部分, 介绍了核医学背景下的 ML。文章阐述了 ML 的历史, 描述了常见的算法, 并举例说明了这些算法在核医学中的应用。第二部分将集中讨论 ML 目前对核医学领域的贡献, 讨论了未来对 ML 的期望和限制, 并对 ML 能做什么和不能做什么进行非常严谨的评估。

参 考 文 献

[1] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev.* 1958;65:386-408.

[2] Silink K. The possibility of designing machines which learn diagnostic: the zerosystems of types and pathotypes in endocrinology. *Act Nerv Super (Praha)*. 1961;3:148-153.

[3] Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge, MA: MIT Press;2016;1-800.

[4] Patterson D. 50 years of computer architecture: from mainframe CPUs to neuralnetwork TPUs. *Proc IEEE Int Solid-State Circuits Conf.* 2018;27-31.

[5] Kelchtermans P, Bittremieux W, De Grave K, et al. Machine learning applications in proteomics research: how the past can boost the future. *Proteomics*. 2014;14:353-366.

[6] VanderPlas J. *Python Data Science Handbook: Essential Tools for Working with Data*. Sebastopol, CA: O'Reilly Media, Inc.;2016;1-548.

[7] Mehta R, Cai K, Kumar N, et al. A lesion-based response prediction model using pretherapy PET/CT image features for Y90 radioembolization to hepatic malignancies. *Technol Cancer Res Treat.* 2017;16:620-629.

[8] Van Weehaeghe D, Ceccarini J, Delva A, Robberecht W, Van Damme P, Van Laere K. Prospective validation of ¹⁸F-FDG brain PET discriminant analysis methods in the diagnosis of amyotrophic lateral sclerosis. *J Nucl Med.* 2016;57:1238-1243.

[9] Ingrisch M, Schöppe F, Paprottka K, et al. Prediction of ⁹⁰Y radioembolization outcome from pretherapeutic factors with random survival forests. *J Nucl Med.* 2018;59:769-773.

[10] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* 1986;32:533-536.

[11] Dechter R. Learning while searching in constraint-satisfaction problems. *Proc 5th Nat Conf on Artificial Intelligence.* 1986;1:178-183.

[12] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput.* 2006;18:1527-1554.

[13] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proc 25th Int Conf on Neural Information Processing Systems.* 2012;1:1097-1105.

[14] Setio AAA, Ciompi F, Litjens G, et al. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Trans Med Imaging.* 2016;35:1160-1169.

[15] Cheng JZ, Ni D, Chou YH, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep.* 2016;6:24454.

[16] Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. *Inf Process Med Imaging.* 2015;24:588-599.

[17] Suk HI, Lee SW, Shen D. Alzheimer's Disease Neuroimaging Initiative. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. *Neuroimage.* 2014;101:569-582.

[18] Li R, Zhang W, Suk HI, et al. Deep learning based imaging data completion for improved brain disease diagnosis. *Med Image Comput Assist Interv.* 2014;17:305-312.

[19] Ciompi F, de Hoop B, van Riel SJ, et al. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. *Med Image Anal.* 2015;26:195-202.

[20] Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging.* 2016;35:1285-1298.

[21] Brosch T, Tam R. Initiative for the Alzheimers Disease Neuroimaging. Manifold learning of brain MRIs by deep learning. *Med Image Comput Assist Interv.* 2013;16:633-640.

[22] Teare P, Fishman M, Benzaquen O, Toledano E, Elnekave E. Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement. *J Digit Imaging.* 2017;30:499-505.

[23] Walker MD, Bradley KM, McGowan DR. Evaluation of principal component analysis-based data-driven respiratory gating for positron emission tomography. *Br J Radiol.* 2018;91:20170793.

[24] Blanc-Durand P, Van Der Gucht A, Verger A, et al. Voxel-based ¹⁸F-FET PET segmentation and automatic clustering of tumor voxels: a significant association with IDH1 mutation status and survival in patients with gliomas. *PLoS One.* 2018;13:e0199379.

[25] Tsujikawa T, Rahman T, Yamamoto M, et al. ¹⁸F-FDG PET radiomics approaches: comparing and clustering features in cervical cancer. *Ann Nucl Med.* 2017;31:678-685.

[26] Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. *AJR.* 2017;208:754-760.

[27] Thrall JH, Li X, Li Q, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol.* 2018;15:504-508.

[28] Vallières M, Zwanenburg A, Badic B, Cheze Le Rest C, Visvikis D, Hatt M. Responsible radiomics research for faster clinical translation. *J Nucl Med.* 2018;59:189-193.