

人工智能深度学习诊治癌症

利用人工智能诊治疾病是人类一个雄心勃勃的计划，而且早就有一些初步结果。如美国国际商业机器公司 (IBM) 2007 年推出的人工智能软件——沃森医生。去年 1 月，时任美国总统奥巴马宣布的“癌症登月计划”，其中一个项目就是让人工智能进行深度学习，以识别癌症。

在癌症诊治领域，人工智能的学习任务有三：通过深度学习在分子层面认识癌症；进行药物筛查；通过大数据分析，提供最佳治疗策略。

现在，具备深度学习功能的机器人医生不但已经进入实践，而且取得了可喜的进展。

■张田勘

人工智能如何诊治癌症

要让人工智能诊治癌症，首先需要人工智能有像人一样感知周围环境的能力，尤其是能感知人的机体状态——什么是正常的机体，什么是异常的机体，甚至是癌变的机体，尤其是只具有微小变化的机体，例如，只有几个细胞发生癌变的极为微小生物环境。

第二步是，人工智能不仅要感知正常和异常机体的不同，还要理解为何有这样的不同，是癌变引起的不同，还是其他疾病引起的不同。最后第三步才是判断和决策，即得出结论，在感知和正确理解的基础上，向医生提供对某个个体检测的结果，是患癌还是没有患癌，抑或是患了其他疾病。

机器学习（算法）是人工智能的一个基本内容，其中，数据的输入、输出、赋值等运算，可以让人工智能对某一问题进行计算分析，从而得出初步结果。这也可以用来对癌症进行诊断和治疗。

加拿大安大略省西方大学的罗根等人研发了一套算法，通过对基因数据的分析得出最可能的有效治疗癌症的方案，并且让该治疗方案变得更加个性化。

研究人员使用了一套含有 40 个基因的数据，这些基因可以在 90% 的乳腺癌中找到。在接受试验的近 350 名癌症病人中，至少都会接受一种化疗药物如紫杉醇或吉西他滨的治疗。之后，研究人员让人工智能对数据展开处理，并找出药物与病人基因之间存在的关系。结果显示，同时接受两种药物治疗的有效率为 84%，只接受紫杉醇的为 82%，只接受吉西他滨的则在 62%-71% 之间。

这就为医生提供了选择更好或最佳治疗方案的决策基础，例如，在上述方案中，医生选择对病人同时使用紫杉醇和吉西他滨，可以达到最高的 84% 的治疗有效率。

2016 年，日本研究人员称，他们开发的人工智能软件能够准确诊断出女性患者所患的罕见类型的白血病，而且，这种软件对肿瘤大数据的提取和分析是其优势之一，它仅需要花费 10 分钟时间，就能够对来自 2000 万个临床肿瘤研究所提供的女性遗传信息进行对比分析，从而做出诊断。

但是，最早开发人应用于医疗领域的人工智能的美国国际商业机器公司 (IBM) 更是走在了前面（见相关链接——从“沃森医生”到“沃森基因组”）。

深度学习的三大任务

2016 年 1 月，美国总统奥巴马宣布了“癌症登月计划”，其中一个项目就是让人工智能进行深度学习，以识别癌症。

在癌症诊治领域，人工智能通过深度学习，要完成三项任务——

首先是认识癌症的分子层面学习，让人工智能了解蛋白质相互作用的机理。

其次是进行临床前的药物筛查。这是一种研发癌症药物的预测模型，在临床试验前进行最大化地药物筛选，为癌症病人提供精准医疗方案。具体而言，是对临床前和临床试验时的癌症数据进行筛选，结合小鼠模型中的新数据，通过反馈循环让实验模型指导计算模型设



计，建立肿瘤药物反应的预测模型。其实，这也是基于特殊数据的学习和分析。

第三是建立人口模型。这就要求人工智能根据不同人群的生活方式、生活环境、所患癌症的种类、不同的医疗体系等，从数百万癌症病人的病历数据中自动分析，从而获取最佳治疗策略。当然，海量病人的数据来自美国国家卫生研究院、美国食品和药品管理局、制药公司和第三方支付机构。

可以看到，美国的“癌症登月计划”中的人工智能深度学习并不包含肿瘤图像的识别，所以人工智能诊治癌症的深度学习在不同的国家有不同的内容。

快速精确解读肿瘤图像

诊断癌症不仅要靠解读癌症特有的基因、分子标记物等，还要认识和判断用各种物理和化学方式拍摄的人体肿瘤图像，这既是人工智能深度学习的内容，又是人工智能帮助人类诊治癌症的一个重要途径，在这个方面，人工智能也已取得了一些进展。

2016 年 8 月，美国休斯顿理工公会医院的研究人员在《癌症》杂志上发表文章称，他们开发的一款人工智能软件在解析乳腺 X 光图片时，比普通医生快 30 倍，其准确率更是高达 99%。这个癌症诊断软件可以直观地将 X 光图片的信息转译成诊断信息，方便医生快速对病人病情作出判断，避免耽误病情。

即便是肿瘤科的专科医生，对 X 光片、CT 和核磁共振成像图片的解读都不会是百分之百的准确。美国疾控中心 and 癌症协会的数据显示，每年美国大约有 1210 万人接受乳腺 X 光片监测，但其中差不多有一半在 X 光片上都有阳性结果，但实际上是假阳性。为此，又迫使大量女性为了求得安心而进一步接受乳腺活组织检查，进行这一检查的人每年有 160 万人左右，其中 20% 的女性根本就没病。这给许多女性及其家庭造成了极大经济和精神负担。

为了改变这种状况，研究人员打算从人工智能着手，来解决 X 光片识别癌症的假阳性问题。理工公会医院的研究人员设计的这个人工智能软件，不但能扫描病人的 X 光影像结果、采集诊断特征，并将乳腺 X 线影像结果与乳腺癌亚型进行关联。此后，医生可使用软件的分析结果来精确预测每个病人是否有罹患乳腺癌的风险。

利用这个人工智能软件，研究人员解读了 500 名乳腺癌病人的乳腺 X 线影像结果和病理组织切片报告，同时还安排了各种相关医学数据，试图迷惑人工智能软件。另一方面，让两名肿瘤科的权威医生进行同样的读片诊断。人工智能用了几个小时就完成了对诊断，但两位乳腺癌专业的权威医生却花了 50-70 小时。

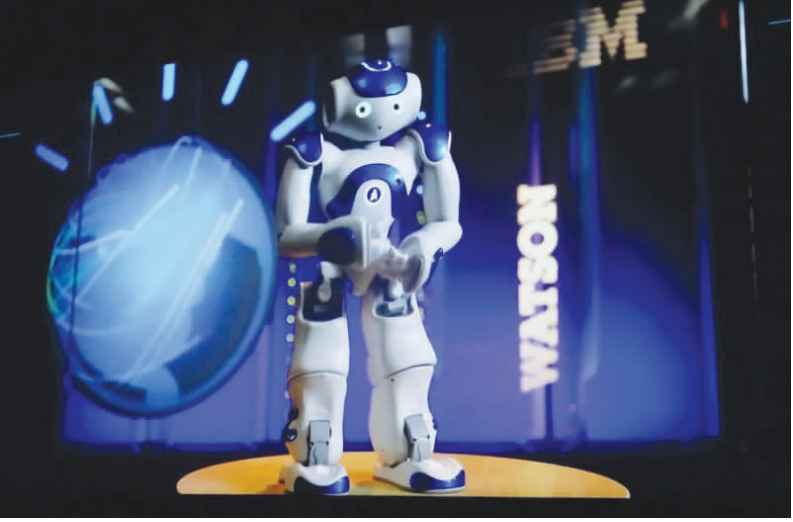
这套软件能够快速和准确诊断乳腺癌的原因也在于，它能在很短时间回顾几百万份记录，通过解读病人的乳腺 X 线影像结果来帮助医生诊断，效率更高。如此，也有望减少不必要的组织活检。

目前，乳腺癌组织活检结果的准确率只有 60%-80%，美国每年浪费在最简单的非侵入性乳腺癌上的资金也高达 3500 万美元。

预计到 2024 年，美国乳腺癌活检设备的市场规模将达 9.11 亿美元。

相关链接

从“沃森医生”到“沃森基因组”



美国 IBM 公司之前推出的人工智能软件——沃森医生，是一个建立在对大数据的检索、使用和算法之上的疾病诊治系统。沃森医生储存了数百万的文档资料，包括字典、百科全书、新闻报道、文学作品及其他可以建立知识库的参考材料。沃森医生的硬件配置可以使它每秒处理 500GB 的数据，相当于 1 秒阅读 100 万本书。

沃森医生在面临一位就诊者的时候，会进行一系列的算法，包括语法语义分析、对各个知识库进行搜索、提取备选答案、对备选答案证据的搜寻、对证据强度的计算和综合等。此外，沃森医生还可以通过询问病人的病症、病史，迅速给出诊断提示和治疗意见。通过这些程序进行诊断，沃森医生的诊断准确率达到 73%。

现在经过多年的改进，沃森医生在对癌症的识别和诊断方面有了突破。最近，IBM 和美国著名的 Illumina 基因公司进行合作，在沃森医生的基础上，专门进行癌症基因组的标准化测序和解读，用以诊断癌症。根据这个目标，IBM 研发了一个新的专门对基因组进行测序和分析的软件，即沃森基因组（相当于专门诊治肿瘤的专科医生），并将这个软件整合到 Illumina 公司的肿瘤测序计划中，这就可以让沃森基因组能够使用 Illumina 公司的

实体肿瘤分析面板。

新的智能软件融合后，沃森基因组可以在短短几分钟内，读取实体肿瘤分析面板生成的遗传信息文件，梳理专业指南、医学文献、临床试验汇编和其他知识来源。然后，系统将生成包含每个基因组改变的注释报告。使用沃森基因组可以大幅减少解释结果所花费的时间。

比较起来，研究人员也可以使用实体肿瘤分析面板进行癌症基因的检测，但是速度很慢。沃森基因组在几分钟内做的事情，研究人员一般需要一个多星期来做。

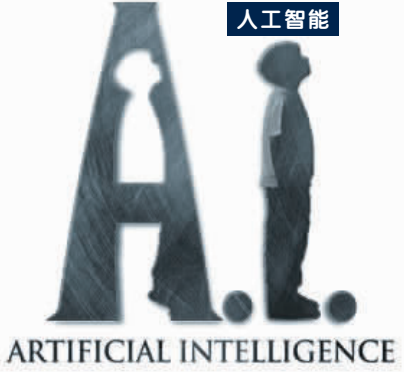
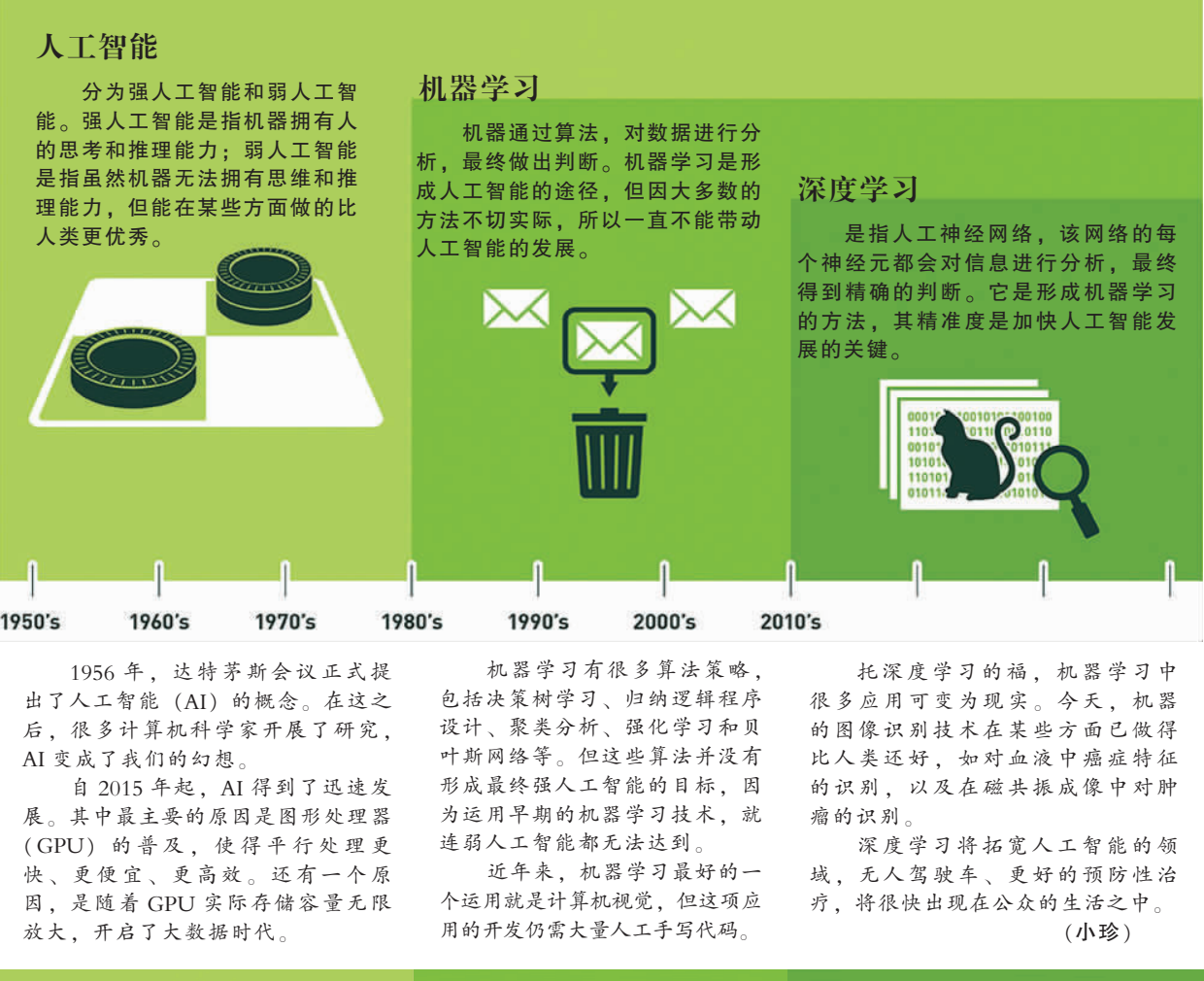
不仅在速度上沃森基因组可以比人类快得多，而且在检测的准确性以及提供治疗癌症的方式上，沃森基因组与临床大夫和肿瘤专家提供的方案基本一致。美国北卡罗来纳大学教堂山分校的夏普尼斯博士研究了 1000 余名癌症患者的数据，发现在 99% 的病例中，沃森基因组提出的治疗建议与分子肿瘤专家团提出的治疗建议相同。此外，在 30% 的肿瘤病例中，沃森基因组还发现了癌症专家遗漏的一些细节。

基于这些结果，研究人员认为，教会人工智能诊治肿瘤大有可为。现在，美国 20 个专注于基因组学和肿瘤学领域的癌症研究所，包括纪念斯隆-凯特林癌症中心和北卡罗来纳大学教堂山分校的肿瘤研究机构，正在进一步培训沃森基因组，以便更快更好地诊治癌症。（田地）

人工智能、机器学习、深度学习之间的区别

去年上半年，谷歌研发的阿尔法狗打败了韩国围棋大师李世石。在这之后，媒体纷纷采用人工智能、机器学习、深度学习这三个词语，来解释阿尔法狗获胜的原因，但它们到底有什么区别呢？

这三者的关系可以看成是一个同心圆。最外层最大的圆是人工智能，是一个最广泛的概念；中间的圆是机器学习；而核心圆是驱动人工智能发展的深度学习。



用人工智能拯救地球

■容沁 编译

2016 年可以说是当之无愧的人工智能 (AI) 年。AI 在大数据、机器学习、计算效率，以及远超人类精确度的算法等方面的创新和优势，给我们的日常生活带来种种便利。和人工智能联手，人类现在可以轻松完成更多的工作。

让 AI 帮我们订购食品、做备忘录或打游戏等，实在是太大材小用了。AI 的超强能力，还能帮助人们处理世界最难最紧急的问题。而其中最基本的一个问题是，如何阻止生物多样性毁灭性地持续下降。

摆在眼前的事实是，物种灭绝的速度很快，比人类灭绝的速度至少要快 1000 倍。而这些物种的灭绝对于人类赖以生存的生态环境来说，会是致命的打击。物种的缺失会严重导致授粉环节停滞，我们的农业系统就会濒临崩溃；生活中必不可少的饮用水与森林、湿地的质量有着紧密的联系，而森林和湿地质量的好坏，又取决于一张复杂的物种关系网络。

还有更让人担忧的。有科学家预计，世界上会有 1000 万物种会灭绝，但他们总共列出其中的 150 万种，目前还未能详细地说出剩余将会灭绝的物种。还有一些科学家预计，灭绝的物种将会达到 5000 多万。按这个预计，《世界自然保护同盟濒危物种红色名录》里列出的濒危动物，仅为估算数字的 1%，甚至更低。

可惜的是，要查清所有可能灭绝的物种，以目前我们的技术，至少还要 500 年！等科学家们弄清时，大多数的物种差不多已经灭绝了。

这个过程之所以这么慢，是由两个因素引起的。一是资金：若想加快查清灭绝物种这个过程，希望在 50 年之内完成，那么每年需要花大约 10 亿美元。二是时间：对已有数据的采集、归类、分析需要耗费大量的时间和精力，而要采集新发现濒危物种的数据，则需要耗费更多的时间。

幸运的是，AI 和一些相关的技术不但能帮助我们降低成本，而且可以更高效地解决这个问题。目前，我们的硬件设施成本越来越低廉，可以让我们在地面、动物身上、天上、甚至是太空，有效地部署监控系统。先前的研究表明，我们可以通过一些公式对这些监控系统中的数据进行筛选。这样一来，我们就能和机器一起，辨认观察由远程设备捕捉到的植物、鸟类、鱼类和其他物种。而且，我们每天都在探寻运用这些技术的新方法。比如，微软正在用蚊子类有机物做实验，把蚊子看成小的、自力更生的数据收集设备，通过分析它吸食哪些其他物种，来观察我们的生态系统。

AI 技术还能帮助我们理解土地利用模式。目前，微软和其他团队正试验将高分辨率图像转变成土地覆盖图。这种土地覆盖图会给我们提供一种史无前例的视野，让我们了解那是什么，它又是如何演变的。有了这种土地覆盖图，政府、组织机构和研究人員就能得到更多的信息，从而采取更有力的保护措施。如通过对非洲大草原的象群数量覆盖调查，发现生活在里的大象数量 7 年间下降了 30%。各方融合这些信息后，可以知道何时、何地及如何更有效地去配置保护措施，而这些新的信息又会反馈到 AI 系统，使 AI 变得更加聪明。这就形成了一个良性的学习循环。

现在我们知道，AI 技术能帮助我们制造一个数字化生态仪表盘，让我们能以前所未有的方式，对生态环境进行监督、规范和管理。而考虑到时间紧迫、资源稀缺，我们必须马上对 AI 技术进行深入的目的地研究。我们需要更多的信息，以及更实质性的方法。联合国近日就宣布，要在 2019 年发表《地球生物多样性和生态系统评估》，而上一次的全球评估是在 2004 年。

那么，我们究竟如何将这种设想变为现实呢？目前，我们刚刚迈出第一步——了解了人类和机器能够共同完成保护生物多样性的目标，而要领悟整个过程，需要政府、学术界还有各行业坚持不懈的努力。这必将不是一个简单的探索之旅，但有 AI 对人类的和我们自然系统的支持，一定会让地球拥有一个更加美好和健康的未来。

（原文作者卢卡斯·约帕博士系微软环境科学家）