深度神经网络会产生人这样的智能吗？ | 人工智能专栏

原创 2016-09-21 王培 赛先生





“人工神经网络”和人脑中的“神经网络”只有非常有限的共同点，而严格模拟人脑的神经网络也未必能解开智能之谜。人工智能和人类智能应该是“神似”而非“形似”。

撰文

王培（美国天普大学计算机与信息科学系）

深度学习在近年的进展又一次点燃了各界对人工神经网络的热情。这一技术在图像识别、语音识别、棋类游戏等领域的成效出人意料，而且更多应用领域也正在被开拓出来。“深度学习是否有效”已经不是问题，现在的问题是在哪些问题上有效，尤其是这条研究路线是否是达到通用智能

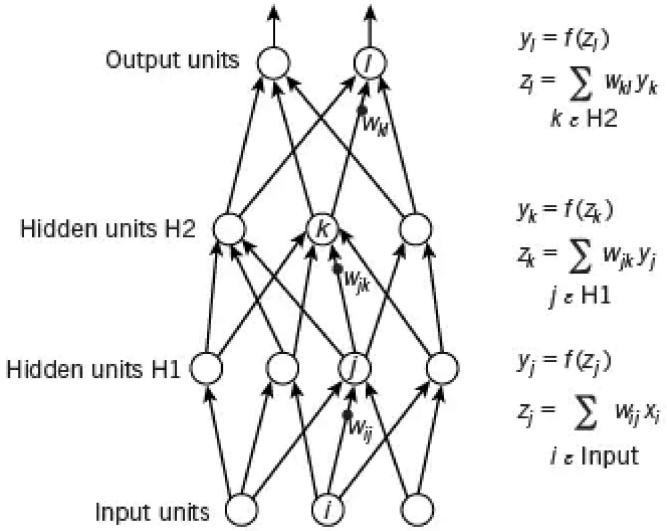
的最佳途径。我在前面几篇短文中涉及到了这个话题，但均未展开谈。关于人工神经网络的工作原理和这项研究的历史沉浮，有关介绍已有很多，这里不再重复。我主要想讨论几个被普遍忽视或误解的概念问题。

此网络非彼网络

在实现“像人一样的智能”的诸多可能途径中（见《当你谈论人工智能时，到底在谈论什么？》），人工神经网络似乎具有天然的合理性和说服力。我们都知道人的智能来自人脑，而人脑是个神经网络，不是吗？

当然没这么简单。所谓“人工神经网络”和人脑中的“神经网络”只有非常有限的共同点，而不同点则要多得多。我们不能仅仅因为它们名称上的相似性就断定它们会有相同的功能。

下图是参考资料[1]中的深度神经网络：



这类网络由若干层组成，每层中的人工神经元与相邻层中的神经元相连接。网络中的底层接受输入信号，顶层生成输出信号，中间层将下层的输出值做加权求和后经一个“激活函数”产生成本层输出值，以供上层之用。这样，每层将一个“向量”（即一串数值）变成另一个向量，而整个网络则代表了一个从输入层到输出层的“向量函数”。这里的输入可以是各种感知信号，中间层代表信号的概括和抽象，而输出则代表系统的认知结果或应对行为。

说这种系统能“学习”，是指在构建网络时，设计者只需选定神经元模型（如激活函数的公式）和网络结构（如一共几层，每层多少神经元）等，而将各个连接上的权值作为待定“参数”。在网络的“训练”过程中，一个“学习算法”根据已知数据对这些参数反复进行调整，直到整个网络达到某种预定的标准为止。这时我们就说这个网络“学会了”如何识别、区分或应对这种输入信号。如我在《机器是如何被骗并骗人的？》中所解释的，这种网络就是个通用函数拟合器。

这种网络结构在三十年前就已经以“PDP模型”的名称广为流传了，只是那时的网络一般只有一个中间层。“深度学习”的贡献就是增加其“深度”，即中间层的数量。更多的中间层使得系统的行为更复杂，这是显而易见的，但以前无法有效的训练这种系统。深度神经网络成功的原因公认有三个：（1）大量的训练数据（否则无法确定那么多参数），（2）改进了的学习算法（否则无法有效利用这些数据），（3）强有力的硬件支持（否则无法足够快地完成计算）。

这种人工神经网络的确是从人脑得到了灵感，但并不是以严格模拟后者为目标的。首先，人工神经元忽略了人脑神经元的很多重要特征，比如说激活的时间模式等。其次，人脑中的神经元有很多种，互相的连接方式也远比上述模型中的要复杂。最后，人脑的学习过程不是通过反复调整权值而“收敛”于一个特定的输入输出关系的，所以不能被简化为函数拟合。

当然，前面提到的人工神经网络只是现有诸多模型之一，其它模型在某些方面可能更像人脑一些。尽管如此，目前的各种人工神经网络和人脑中的神经网络仍有巨大差别。仅仅因为一个计算机中用了这样一个网络，就期待它能像人脑一样工作，这未免太天真了。实际上这种网络的设计者主要关心的是其实用性，并非是否像人脑。

智能难以一“网”打尽

试图严格模拟人的神经网络的另有人在，只不过他们的目的主要不是为了实现人工智能，而只是想了解人脑的工作过程。有些研究者已经把“全脑仿真”（whole brain emulation）作为目标了。这项工作当然是很有意义的，因为“模拟”是我们认识一个对象或过程的有效手段。一般说来，只要我们能精确地描述一个过程，我们就可能编写计算机程序来模拟它。但即使我们真能精确地模拟人脑中神经网络的工作过程，这也离完整地实现人类智能有相当远的距离。

首先，人脑不仅仅是个对电信号进行处理的神经网络。比如说神经递质在神经元之间担当了“信使”的作用，而其中的活动是化学过程。甚至人脑中的生物过程和物理过程都可能对思维产生影响，比如脑供血不足和剧烈运动后的眩晕现象。

即使在理论上说人脑中的上述过程也都能模拟，这事也还没有完。近年来认知科学越来越强调躯体在思维中的作用（所谓“embodied cognition”，即“具身认知”），就是说大脑之外的躯体部分（如外周神经系统、感觉器官、运动器官等等）也在思维活动中扮演着不可或缺的角色。比如说，很多抽象概念的意义都是基于躯体活动的(如“接受”批评、“推动”发展、“提高”觉悟等)，那么要完整理解这些概念，一个系统大概也需要真能“接”、“推”、“提”才行。

“模拟派”的追随者可能会说人体也是可以用软硬件模拟的。这“在原则上”自然是正确的，但如果神经网络之外的部分对实现智能是必须的，那么单靠人工神经网络来完整再现智能大概就是不可能的了。为实现“像人一样的智能”，我们需要的不再是“人工神经网络”，而是包括这样一个网络的“人造人”。

就算一个“人造人”被制成了，它大概也不会是所有人心目中的“人工智能”。对那些以“图灵测试”为智能标准的人而言，“智能”意味着在外部行为上和人不可区分。大部分人都同意完全靠预先设计所有可能的答案来通过这个测试是不大现实的，而通过“学习”才有希望。但这就意

味着只有类人的躯体还不够，系统还需要类人的经验。这就不再仅仅是个技术问题了。因为我们的经验中的很大一部分是社会经验，只有当人造人完全被当作人来对待时，它才能得到人类经验，进而获得人类行为。而这将会在伦理、法律、政治、社会等领域造成大量问题。

“忠实复制”不是好主意

上面讨论的是用人工神经网络实现像人一样的通用智能的**可能性**，其结论是：远没有看上去那么有希望，尽管不是完全不可能。但这条路线比可能性更大的问题是其**合理性**和**必要性**。

人工智能的基本理论预设是把人类智能看作“智能”的一种形式，而试图在计算机中实现其另一种形式。根据这个看法，“人工智能”和“人类智能”不是在细节上完全一样，而是在某个抽象描述中体现着同一个“智能”。如果智能所需的某个机制在计算机里有更好的实现方式，那我们没必要用人脑的办法。这方面的一个例子就是四则运算。

当我们为一个对象或过程建立模型时，我们总是希望这个模型越简单越好。只有当我们可以在忽略了大量细节的情况下仍然可以准确地刻画一个过程，我们才算是真正理解了它。因此，如果最后发现我们只有在严格复制人脑、人体及人类经验的情况下才能再现智能，那么人工智能应当算是失败了，而非成功了，因为这说明智能只有一种存在方式，而“智能”和“人类智能”其实是一回事。

出于这种考虑，很多人工智能研究者有意识地和人脑的细节保持距离。在从人脑的工作方式中得到灵感的同时，他们会考虑在计算机里是否有更简单的办法来实现同样的功能。计算机毕竟不是个生物体，所以没必要模仿人脑的那些纯生物特征。同理，对深度神经网络的研究者来说，他们的网络在某些方面不像人脑，这不是个问题，只要这种差别不带来功能缺失就行。

这个问题是所有走“仿生”路线的技术都要面对的，包括近来大热的“类脑智能”、“类脑计算”等。在有关讨论中，一些人只是强调“像人脑那样”的可能的好处，但完全不提这种模仿的**限度**。只有当我们能清楚地说明哪些东西不用模仿时，我们才算真正说清了哪些东西需要模仿。只是说“人脑是这样的”尚不能成为“计算机也必须这样”的充分理由。

这里一个常被提到的例子就是飞机和鸟的关系。飞机的初始设计的确借鉴了鸟类，但显然不是越像鸟越好。这里自然有可能性的考虑，但更重要的是要飞机完全像鸟既不合理也无必要。我们当前的课题也同样。严格说来，“人工神经网络”、“人工大脑”、“人造人”和“人工智能”各是不同的研究目标，各有各的价值和意义。尽管它们之间有联系，仍然不能混为一谈，因为设计目标和最佳实现方法均不同。

如何“取长补短”

在人工智能历史上，以神经网络为代表的“联结主义”和以推理系统为代表的“符号主义”已经竞争了多年，彼此地位的消长也经历了几次反转了，这有点像光的“粒子说”与“波动说”的斗争史。时至今日，大部分人都会承认这两个传统各有长处和短处。在这种情形下，把二者结合起来就是个自然的想法了。

问题是怎么“结合”。一个常见的办法是把两个基于不同传统的子系统整合在一起，让它们各自完成自己擅长的工作，并彼此协作来完成整个任务。尽管具体做法不同，在这个方向上的探索成果已经不少了。

我自己的方案是把这两个传统以一种更密切的方式统一起来。如我在参考资料[2]中所论证的，我认为联结主义的哲学假定更正确，但技术手段太单一；与其相反，符号主义的技术手段更丰富，但哲学假定陈旧呆板。其结果就是我的“纳思”（见《你这是什么逻辑?》）在理论预设和宏观战略上更接近于联结主义传统，如假定知识和资源不足，容忍多种不确定性，通过学习获得领域知识和技能等；而纳思在描述框架和微观战术上更接近于符号主义传统，如用语言表达知识，靠概念组织经验，依逻辑进行推理等。纳思不是把这两个传统“嫁接”在一起，而是把它们“杂交”成了一个全新的“神经（网络式的）逻辑”。

像我在《“意思”是什么意思？》中所介绍的，纳思也可以被看成一个网络，但它和人脑的相似性不是在神经元层面上，而是在概念层面上。这个网络的运行方式和现有的人工神经网络根本不同，尽管其中的概念关系也支持多层抽象。那样的话，纳思到底是个逻辑还是个网络呢？都是。对象只有一个，但对它的描述可以使用不同的术语，以捕捉其不同的侧面。这和光的波-粒二相性不乏类似之处。

总而言之，我不认为深度神经网络会产生通用智能，或是通向这一目标的有效途径，尽管这种技术有巨大实用价值，并可以为通用智能的研究贡献想法。人工智能的确要“以脑为师”，但若是亦步亦趋，恐怕会落个邯郸学步的下场。像我在《人工智能：何为“智”？》中所说的，人工智能和人类智能应该是“神似”（体现在基本原理上）而非“形似”（体现在结构或行为上）。

参考资料

[1] Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Deep Learning. Nature 521: 436-444, 2015

[2] Pei Wang. Artificial general intelligence and classical neural network. Proceedings of the IEEE International Conference on Granular Computing, 130-135, May 2006

了解作者信息和他在赛先生上已发表的文章，请点击页面左下角的“阅读原文”访问他的主页。

投稿、授权等请联系：iscientists@126.com

您可回复"年份+月份"（如201510），获取指定年月文章，或返回主页点击子菜单获取或搜索往期文章。



赛先生由百人传媒投资和创办，文小刚、刘克峰、颜宁三位国际著名科学家担任主编，告诉你正在发生的科学。上帝忘了给我们翅膀，于是，科学家带领我们飞翔。

微信号：iscientists

▲

长按图片识别二维码关注我们

点击“阅读原文”购买科学好书!

Read more