

连接组、深度学习以及未来五年的 AI

连接组这个词从**基因组**（genome）衍变而来，用来表示人类全部神经元及其连接结构。连接组学主要是通过分析神经元之间的连接和组织方式来达到分析大脑的运行机制这一终极目的的一门学科。**深度学习**是机器学习的一个分支，是受人类大脑神经元结构启发的学习算法。深度学习还能从神经科学汲取哪些营养？未来五年内的 AI 能达到什么程度？我们应该从什么方向进行探索？本文没有太多所谓的干货，就是想和广大人工智能爱好者异想天开的畅想一下人工智能的未来。

一、神经元

生物神经元，又称神经细胞，是构成神经系统结构和功能的基本单位。它由细胞体和细胞突起构成。神经元的突起是神经元胞体的延伸部分，由于形态结构和功能的不同，可分为树突和轴突。典型的神经元结构如图1-1所示。生物神经元可以通过刺激产生冲动、传导冲动；并可以合成化学物质（神经递质、神经激素等），通过他的轴突输送到特定部位并释放，从而产生一个反应。

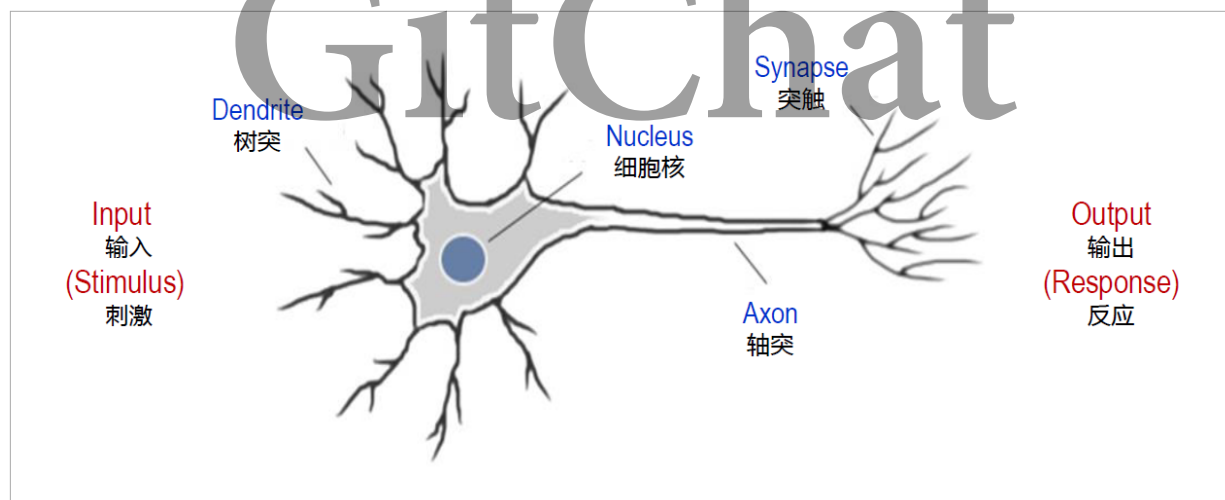


图1-1 生物神经元的结构

第一个正式的人工神经网络模型是由沃伦·麦卡洛克（Warren Macculloch）和沃尔特·皮茨（Walter Pitts）于1943年提出的。这个模型看起来很像组成计算机的逻辑门。人工神经元示意图如图1-2所示。人工神经元本质上是一组向量的加权求和，并通过激活函数产生一个输出。单个的神经元模型无法从数据中学习。

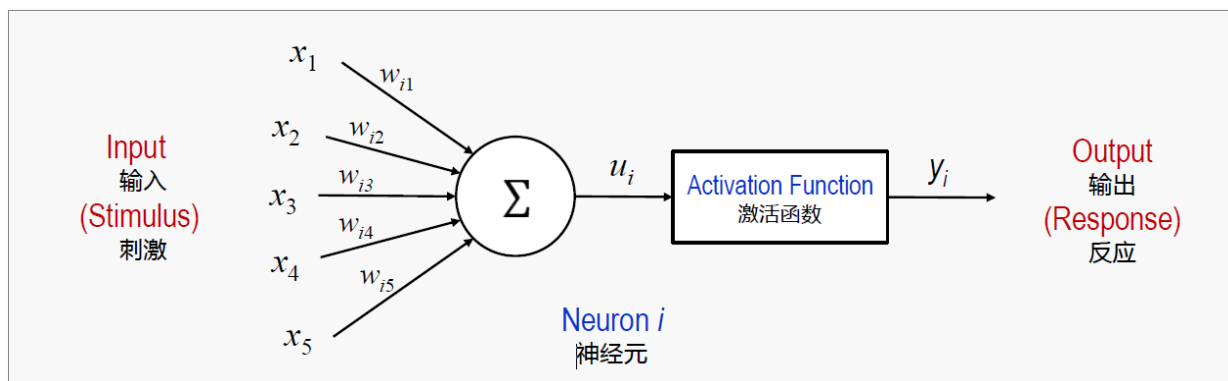


图1-2 人工神经元

人工神经元和生物神经元有两个相似点。第一，他们都有输入（刺激）和输出（反应）；第二，单个神经元都不能表现出复杂的智能。

二、神经网络与连接组

神经网络在人工智能领域主要指人工神经网络。在这里我想先介绍一下生物神经网络。生物神经网络（Biological Neural Networks）一般指生物的大脑神经元，细胞，触点等组成的网络，用于产生生物的意识，帮助生物进行思考和行动。我在读《连接组：造就独一无二的你》这本书时，我对书里一张恒河猴的视觉皮层连接图有很深的印象（见图2-1），当时就在想生物的神经网络连接和人工神经网络有什么关系吗？我们还能不能从生物神经网络里汲取灵感来研发更好的人工智能？这两个问题我会在本文的后续部分给出我的思考。

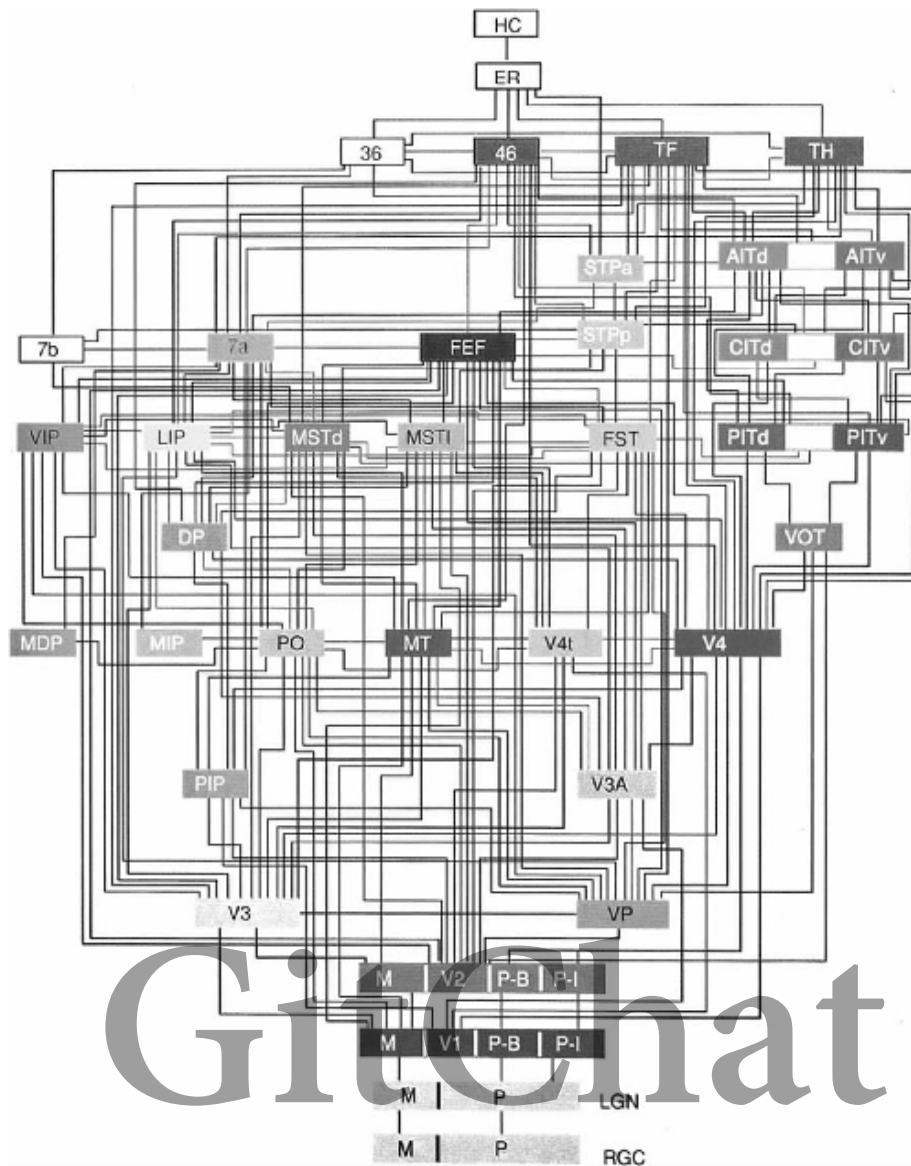


图2-1 恒河猴的视觉皮层连接图

连接组在人的一生当中是变化着的，这种变化在胎儿和婴儿时期是最快的。连接组的变化一般包括重新赋权、重新连接、重新连线和重新生成。这四个“重新”伴随人的一生，只是在不同阶段会出现以某个重新最为显著。这四个重新在提高“正常的”大脑和治疗疾病或受损大脑的过程中，扮演着非常重要的角色。认识这四个“重新”的全部潜能，可以说是神经科学最重要的目标。另外，这四个“重新”的潜能并不是固定的。我们观察到大脑在受伤后，轴突的生长会增强。除此之外，现在已知新皮层受损后会吸引新生的神经元，使它们迁移到伤处，这也是“没有新生神经元”的又一个反例。这些由受伤产生的作用是由某些分子作为媒介的，目前还正在研究。理论上说，应该能通过人工方法，通过操作某些分子来促进这四个“重新”。这是基因向连接组施加影响的方式，也应该是未来药物的作用方式。不过，这四个“重新”还受到经历的引导，所以更好的控制方法是在进行分子操作的同时配合适当的强化训练。

现在让我们再来看看人工神经网络的发展。上文提过1943年出现了第一个人工神经元，但该模型过于简单。1957年，罗森勃拉特（Rosenblatt）以M-P模型为基础，提出了感知机（Perceptron）模型。感知机模型具有现代神经网络的基本原则，并且它的结构非常符合神经生理学。这是一个具有连续可调权值矢量的MP神经网络模型，经过训练可以

达到对一定的输入向量模式进行分类和识别的目的，它虽然比较简单，却是第一个真正意义上的神经网络。单层感知机的局限性是不能解决异或问题，这个缺陷被人工智能创始人之一的明斯基研究发现，并于1969年出版了一本《感知机》的书来论述人工神经网络的局限性。由于明斯基在人工智能领域有相当大的权威，他的这本书给当时人工神经网络的研究带来沉重的打击。开始了神经网络发展史上长达10年的低潮期。1983年，Kirkpatrick等人认识到模拟退火算法可用于NP完全组合优化问题的求解，这种模拟高温物体退火过程来找寻全局最优解的方法最早由Metropolis等人1953年提出的。1984年，Hinton与年轻学者Sejnowski等合作提出了大规模并行网络学习机，并明确提出隐单元的概念，这种学习机后来被称为Boltzmann机。Hinton和Sejnowsky利用统计物理学的感念和方法，首次提出的多层网络的学习算法，称为Boltzmann 机模型。隐藏层概念的提出对神经网络是一个重大突破。至此，多层网络结构成为了神经网络的主流，多层网络可以成功异或问题和其它更复杂的模式识别问题。

三、深度学习

2006年，Hinton在《Science》期刊上发表了论文《用神经网络降低数据维数》，首次提出了“深度信念网络”的概念。与传统的训练方式不同，“深度信念网络”有一个“预训练”（pre-training）的过程，这可以方便的让神经网络中的权值找到一个接近最优解的值，之后再使用“微调”（fine-tuning）技术来对整个网络进行优化训练。这两个技术的运用大幅度减少了训练多层神经网络的时间。他给多层神经网络相关的学习方法赋予了一个新名词-“深度学习”。

尽管 2006 年以后深度学习就开始引起关注，但震撼性的影响发生在 6 年后。在 2012 年 ImageNet 图像识别竞赛上，Hinton 的两个学生采用深度学习的识别结果，准确率超过第二名东京大学10%以上，而第二到第四名都采用传统计算机图像识别方法进行分类，他们之间准确率的差别不超过1%。也就是说，采用深度学习，把图像识别的准确率进步一下子提高了一个数量级。2012 年 10 月，在意大利佛罗伦萨的研讨会上，竞赛组织者李飞飞宣布了这一压倒性的结果，在计算机视觉领域产生了极大的震动，并迅速波及到整个 AI 界和产业界。深度学习的热潮从此掀起，一波接一波向前迅猛推进，不断进入一个又一个领域并连战连捷，势如破竹，形成今天锐不可挡的AI狂潮。

深度学习取得了如此大的成功，是不是预示着深度学习是实现人工智能的最佳技术路线？深度学习又和人类大脑有多大的相似性？关于第一个问题，我认为深度学习是目前实现图像识别、语音识别类任务的最好技术路线，至于自然语言处理、商业趋势的分析预测等场景虽然都取得了不错的成绩，但要取得更好的效果往往需要领域的强力支持。关于深度学习和人类大脑思维过程相似性问题，答案很明确就是相似性不大。比如人类大脑连接组的四个重新只是在深度学习算法在训练阶段有一个粗略的模拟，人脑非常擅长的小数据学习目前在深度学习领域还没有取得太多的进展。黑箱子问题也是深度学习饱受诟病的一个缺点。不过人脑和深度学习有一个共同点，就是目前对于人类来讲他们都具有不可解释性，都是黑箱子。破解深度学习的黑箱和破解人脑的黑箱，在我眼里都是非常有意思的事情。这两件事应该是相互促进，共同发展的。

2017年，深度学习最大的发展恐怕就是Hinton老爷子提出的Capsule网络。Capsule 是一组神经元，其活动向量（activity vector）表示特定实体类型的实例化参数，如对象或对象部分。我们使用活动向量的长度表征实体存在的概率，向量方向表示实例化参数。同

一水平的活跃 capsule 通过变换矩阵对更高级别的 capsule 的实例化参数进行预测。当多个预测一致时，更高级别的 capsule 变得活跃。我们展示了判别式训练的多层 capsule 系统在 MNIST 数据集上达到了顶尖的性能，比识别高度重叠数字的卷积网络的性能优越很多。这个Capsule网络可以看做是对人类多层视觉系统的模拟。从这一点可以看出，人类对自身大脑的解密虽然还很初级，但是每一个在神经科学方面的进展都有可能对人工智能的研发提供灵感，从而实现一个小飞跃。

2018年初，Gary Marcus发表了对深度学习的系统性批判论文。文中列举了深度学习的十大问题。（1）目前深度学习需要大量数据；（2）深度学习目前还是太表浅，没有足够的能力进行迁移；（3）迄今深度学习没有自然方式来处理层级架构；（4）迄今为止的深度学习无法进行开放推理；（5）迄今为止的深度学习不够透明；（6）迄今为止，深度学习并没有很好地与先验知识相结合；（7）到目前为止，深度学习还不能从根本上区分因果关系和相关关系；（8）深度学习假设世界是大体稳定的，采用的方式可能是概率的；（9）到目前为止，深度学习只是一种良好的近似，其答案并不完全可信；（10）到目前为止，深度学习还难以在工程中使用。这十大问题更多的是从实现通用人工只能的角度来论述的，但是在工业界大部分人可能并不关注是否能实现通用人工智能。在笔者看来目前深度学习还是AI算法中最强大，最有力的。不过我们也应该在更多方向上进行探索。

四、人工智能的未来

谈到人工智能的未来，就避不开两个话题——量子计算和类脑计算。量子计算如果能实现那就是非常强大的计算能力，但是如果不了解大脑的运行机制那么就算有再强大的计算能力也很难实现通用人工智能。不过量子计算强大的计算能力可以帮我们解析出完整的人类大脑连接组，在了解人类大脑连接组的情况下再去模拟大脑，进行类脑计算就相对靠谱得多。要想模拟大脑或者神经系统，连接组和神经元类型这两个要素缺一不可。不过，神经元类型的模型所包含的信息，要远远少于连接组，因为大多数科学家都相信，神经元类型的数量要远远少于神经元的数量。可以把神经元类型的数量想象成电子设备的零件种类数量。电子设备只由很少的几种零件组成，比如电阻、电容、晶体管等。所以一台收音机的功能，主要是由它的电路图决定的。人脑的零件清单比这长一些，所以需要花很多年的努力，去为人脑的每一种神经元类型建模。但是这个零件清单的条目，毕竟还是远远少于零件的总数。这就是为什么零件的组织结构才是最重要的。欧战的蓝脑计划就是用计算机模拟人类大脑，虽然进展缓慢，但也对神经元的活动有了一些模拟，只是还没有发现这个模型表现出任何智能。现在的类脑计算在没有搞清楚人脑的连接组和神经元模型的情况下也在进行积极探索，因为搞计算机的人已经等不及神经科学发展的步伐。在未来五年内人类脑科学可能都无法给人工智能太多理论指导，在这种情况下我们构建人工智能系统除了参考脑科学给的一点启发外，还应该重视数学工具的应用，现在的深度学习算法更多的是一种统计模型。

那么既然是谈人工智能的未来，那么有一个问题我在这里也向回答一下。那就是人工智能会不会毁灭人类，在这里我给出的答案是：“不会，你是不是科幻电影看多了”。至少五年内还看不到人工智能毁灭人类的可能性。前几天，LeCun（CNN发明人）飙脏话批“世界首个机器人公民”Sophia，可见他也认为现在的人工智能还只是一个工具，远没有达到人类智能这个级别。至于，遥远的未来则是一切皆有可能。不过，按现在物理学的观点，地球、宇宙也终有毁灭的一天。能不能在地球毁灭前，创造出超过人类的人工

智能还是未知数。想到科幻小说《三体》的情景，人类还没实现可靠的人工智能前就遭受了三体人的入侵，在地球舰队与三体舰队的末日之战中，人类舰队的计算机系统还远远达不到“水滴”那样的强大自主决策能力。在未来五年内，人工智能从业者，先不要想着人工智能会不会毁灭人类，先在更多领域研发出靠谱的，可以部分代替人类工作的人工智能系统。这个理想就需要所有人工智能学者、工程师共同努力来实现。

GitChat