·科学评论·

两个黑箱问题
——深度神经网络和脑神经网络

谷歌DeepMind团队的研究将认知心理学(cognitive psychology)和深度神经网络(deep neural network,DNN)结合在一起,并发现两者在小样本词汇学习(从一个示例中猜出一个单词的意思,one-shot word learning)过程中具有相似之处——DNN和脑神经网络都具有形状偏好(shape bias),相关研究论文《Cognitive psychology for deep neural networks: A shape bias case study》被国际机器学习学会(IMLS)收录[1]。深度学习和神经科学的学科交叉对未来的发展方向具有很强的指导意义,这项研究也为DNN的机理以及脑神经网络的理论基础研究提供了一种可行的解决思路,而真正破解DNN和脑神经科学的黑箱问题依然任重道远。

两个黑箱问题

DNN在很多复杂任务上取得了前所未有的进展,如人脸识别、围棋以及Atari游戏,但是其解决方案也远远超出了我们的理解范围,成为了一个名副其实的黑箱。这种黑箱模型在很大程度上能够简单便捷地解决许多实际问题,但是从科学研究和实际应用的角度上看,我们需要理解并改变世界,只有破解DNN的黑箱问题之后,才能更好地运用并改善该模型,同时提升实际应用中的可靠性。

在这种黑箱问题背景下,建立更加优秀而且可理解的神经网络系统成为一个热门的研究方向。许多研究者都认可:DNN的图像处理模型与动物的视觉处理通路具有相似之处,动物的视觉处理通路存在感受野、方向选择性以及分级处理等特性,而模仿动物视觉信息处理的DNN也具有类似的特性;此外,DNN的快速发展也促进了视觉神经科学方面的研究,形成了有效的正反馈;从更加宏观的角度来看,感受野、方向选择性、分级处理这些特性也会使DNN和动物视觉信息处理在更高层次上具有一定的相似性。这种共同点以及两者相互促进的机制正是开展DNN黑箱问题与认知心理学的交叉研究的基础。破解DNN黑箱问题,首当其冲的是为该问题建立一套完整的问题描述方法及实验研究方法;由于DNN与脑神经网络存在一定的共性,因此DNN黑箱问题与脑神经科学的交叉研究正是一种行之有效的方法。尽管脑神经网络是一个更加庞大的黑箱,但是人们对于大脑的解码走在DNN黑箱问题的前面,对大脑的研究已经建立了一套系统的研究方法,这套研究方法也可以作为DNN黑箱问题的基础。

DNN与脑神经网络的异同点

认知心理学是研究脑神经网络的一门重要学科,认识心理学中一个经典的案例是考察儿童如何识别和标识物体,探索儿童如何从一个示例中猜出一个单词的意义;认知心理学的研究成果表明,儿童会通过采用归纳偏好来消除许多不正确的推理,而且形状偏好强于颜色偏好。在人工智能中,能够进行小样本学习的深度神经网络(matching network),凭借一个孤立样本,就取得了ImageNet图像分类任务中目前最佳性能,而且该模型也更倾向于形状偏好。小样本学习的能力和形状偏好特性是DNN与脑神经网络的共性,但是DNN和大脑终究就是两个差异很大的模型,一点小的共性难以弥补两者之间的鸿沟。

首先,DNN和大脑的拓扑结构有很大的差别,DNN往往具有非常规则的连接,而真实的脑神经网络之间的连接极其复杂,而且不同物种的脑神经网络连接也会有所不同;如大小鼠、猫和猴存在感受野以及视觉信息分级处理的机制,但是大小鼠没有功能柱,而猫和猴存在功能柱(具有相同感受野并具有相同功能的视皮层神经元,在垂直于皮层表面的方向上呈柱状分布);猫和狗的视锥细胞非常少,对颜色不敏感,猴和树鼩则具有丰富的颜色视觉;青蛙和兔子这类的低等动物的视觉信息处理系统对运动的物体非常敏感;每一类生物在漫长的进化过程中,都已经形成了最适应其生存环境的形态结构,而目前DNN模型的拓扑结构与任何一种生物的神经网络结构都相去甚远。此外,DNN的计算方式与大脑的信息处理方式也有很大不同,DNN一般都是确定性的数学模型,给定输入之后,按照给定的计算流程,所有的中间变量以及最后的输出都是确定的;对于大脑来说,给定一个输入,会得到一个确定的输出,但是中间变量不是确定的(即每次看到一个苹果时,大脑会认出这是一个苹果,但是每次都只观察视觉信息处理过程中很小的一个神经环路时,这个环路的状态是变化的,而这种变化却不影响最终的输出)。从输入到输出,DNN只有一条确定的路,而大脑每一次都走了一条不同的路,这就是DNN的确定性与脑神经网络的不确定性之间的矛盾。

加强深度学习与脑科学交叉研究

DeepMind团队认为这项形状偏好的研究表明:认知心理学工具具有揭示DNN隐藏计算过程的能力,同时能够提供一个人类词汇学习的计算模型,对此我们持一定的怀疑态度。首先,DNN和大脑的拓扑结构具有很大的差异,即使二者具有感受野、方向选择性、信息分级处理、以及形状偏好等形同特性,也并不能说明二者的计算过程有多少相似性;此外,学习和记忆也是认知心理学中亟待解决的黑箱问题,在得到透彻理解之前,并不能单方面地成为破解DNN的黑箱问题的武器。值得肯定的是,深度学习与脑科学的交叉研究是未来必然的发展趋势;在2016年,《Nature Neuroscience》上的一篇文章介绍了一种对更高层的视觉皮层的神经活动进行建模的目标驱动分层卷积神经网络(Goal-driven hierarchical convolutional neural networks,HCNNs),该研究表明目标驱动的HCNNs能够揭示视觉皮层信息处理过程形成和发展的机理[2];神经科学和认识心理学并不是单纯的实验科学,只有基于DNN这一类有效的理论分析方法,才能将二者推向更高的层次;从另一个角度看,随着对脑神经网络认识的深入,人们能够发掘出脑神经网络更多的特性,这些特性很有可能成为不断理解与完善DNN模型的切入点,成为破解DNN黑箱问题的可行途径。

参考文献

[1] Ritter S, Barrett D G T, Santoro A, et al. Cognitive psychology for deep neural networks: A shape bias case study[C/OL]//Proceedings of the 34 th International Conference on Machine Learning. [2017-06-29]. https://arxiv.org/pdf/1706.08606.pdf.

[2] Yamins D L K, Dicarlo J J. Using goal-driven deep learning models to understand sensory cortex[J]. Nature Neuroscience, 2016, 19(3): 356-365.

文/袁培江,苏峰

作者简介:袁培江,北京航空航天大学机械工程及自动化学院,副教授;苏峰,北京航空航天大学机械工程及自动化学院,硕士研究生。

(责任编辑　刘志远)