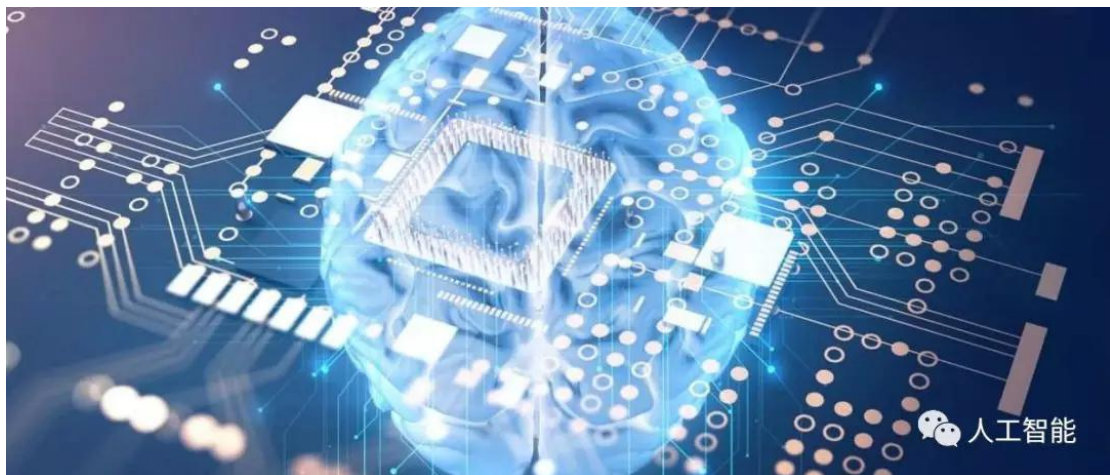


麻省理工团队能将被 AI 优化过的 AI 设计过程加速 240 倍！

人工智能4月9日

自 2017 年以来，人工智能（AI）研究人员一直用 AI 神经网络来设计更好、更快的 AI 神经网络。迄今为止，学界一直在追求用人工智能促进人工智更好地发展。这主要是因为，这种方法能达到数万小时的 GPU 时间。 如果真可以实现这一点，那么设计 AI 应用程序时，可能会更快、更简单。



今年 5 月，麻省理工学院的一个研究团队将展示一种名为“神经架构搜索”（“neural architecture search”）的算法，该算法可以使被 AI 优化过的 AI 设计过程加速 240 多倍。 这有利于人工智能更快、更准，并且在实际范围内广泛应用于图像识别算法和其他相关应用。

自 2017 年以来，人工智能（AI）研究人员一直用 AI 神经网络来设计更好、更快的 AI 神经网络。迄今为止，学界一直在追求用人工智能促进人工智更好地发展。这主要是因为，这种方法能达到数万小时的 GPU 时间。如果真可以实现这一点，那么设计 AI 应用程序时，可能会更快、更简单。

麻省理工学院电子工程和计算机科学助理教授韩松说：“在模型大小、应用延迟、准确性和模型容量之间存在各种权衡。把所有的东西权衡起来就会产生一个巨大的设计空间。以前人们设计的神经网络采用的是启发式算法。神经架构搜索的目的在于减轻算法的工作量，将基于人类启发式的探索转化为算法自主学习的，更符合 AI 的设计逻辑。就像 AI 可以学习下围棋一样，去学习如何设计一个神经网络。”

就像人工智能曾战胜李世石与柯洁，甚至可以教这些围棋国手新的围棋路数，通过 AI 来优化 AI 可以提供新的 AI 设计思路。

麻省理工用的 AI 算法类型是一种卷积神经网络（CNN）。（CNN 通常是用于图像识别的首选神经网络。除此之外，它还被应用于自然语言处理和药物研发。）

韩松说，这种被 AI 优化过的 CNN 的图像分类速度将是其他 AI 的 1.8 倍。

他补充说团队能将 CNN 提速至此，主要缘于以下三点：

首先，他们减少了运行神经架构搜索的 GPU 内存负载。标准神经架构搜索可以同时检查网络中神经层之间所有可能的连接。而韩松团队的算法每次只在 CPU 的内存中保存一条路径。这个技巧可以只用十分之一的内存完成对所有参数空间的搜索，从而使搜索能够覆盖更多的网络配置而不会耗尽芯片上的空间。

第二，该团队通过删减掉冗余的神经网络来精简路径，这样明显加快了神经网络搜索进程。通常，神经架构搜索只丢弃单个“神经元”，因为它会修剪掉所有次优的神经网络连接。

第三，这个算法使神经网络搜索意识到 AI 系统在正在运行的硬件中可能的延迟时间，不管这个硬件是 GPU 或 CPU。

韩松说，令人惊讶的是，通过新的算法，我们发现，传统 AI 中某些图像识别神经网络是错的。从某种意义上说，当前大家已经在用 GPU 运行算法了，但有些做 AI 的人的设计思维仍然处于 CPU 时代。

CNN 在其图像识别算法中使用的是 3×3 、 5×5 或 7×7 像素的正方形网格滤波器，不过此前，大家很少使用 7×7 像素的滤波器，因为人们一直认为运行多层 3×3 像素的滤波器比单个 7×7 像素的滤波器更快。

然而，被 AI 优化的 AI 却使用了不少 7×7 像素的滤波器。韩松表示，当前用于人工智能计算的硬件中，GPU 仍然占主导地位。他说：“GPU 本身有很大的并行性，而且大型内核调用比调用几个小内核调用效率更高，因此 7×7 层的滤波器更适用于 GPU。”