AI能够帮助矩阵乘法更加高效

SY2303526 杨和鹭

在众多的算法中，矩阵起着至关重要的作用，一个深度网络的全部参数都可以以矩阵的方式进行运算，包括根据输入逐层计算最终层的输出值，进行反向迭代做梯度下降等，这些都会用到复杂的矩阵运算。

涉及到高维度的特征运算，海量的数据处理，以及GPU的图像渲染技术，矩阵无疑是最能够使算法高效运转的利器。我们生活中处处隐藏着矩阵相乘，如智能手机中的识别图像处理、语音命令、3A大作中精致的画面。然而令人诧异的是：我们所用的标准矩阵计算方式（即行乘列）并不是最高效的矩阵计算手段，如果能找到一种方法提高矩阵的计算效率，那么在大算力、大数据的时代这一点点效率的提升都会节省大量的人力物力。因此遍布世界各地的公司都愿意花费大量的时间和金钱开发计算硬件以更有效地解决矩阵相乘。

如下图所示：对于二维矩阵，我们的计算方式（左图）相比另一种矩阵相乘的Strassen计算方法（右图）就多做了一次运算，为了达到减少运算次数，代价是右侧的计算公式要更为复杂，但对于计算机的运算方式来说，右侧的矩阵相乘计算方法极大地提升了运算速度。

因此，之后数十年，研究者都在研究更大的矩阵，甚至找到 3x3 矩阵相乘的高效方法，都还没有解决。



DeepMind 的最新研究探讨了现代 AI 技术如何推动新矩阵乘法算法的自动发现。基于人类直觉（human intuition）的进步，对于更大的矩阵来说，AlphaTensor 发现的算法更为有效。该研究表明 AI 设计的算法优于人类设计的算法，这是算法发现领域向前迈出的重要一步。AlphaTensor 发现的算法还首次在一个有限域中改进了 Strassen 的二阶算法。这些用于小矩阵相乘的算法可以当做原语来乘以任意大小的更大矩阵。利用这些特性，DeepMind把成果部署在给定计算硬件（如DSP计算加速模块、显卡 Nvidia V100 GPU）上运行，这些算法在相同硬件上进行大矩阵相乘的速度比常用算法快了 10-20%。

矩阵乘法是神经网络、图论、滤波算法等的核心组成部分，对于这种矩阵运算效率的提升将会助力于相关产业的进一步发展。