

---

# IBM新芯片加速人工智能

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/24670.html>

**本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！**

IBM新芯片加速人工智能。美国IBM公司研究人员开发了一种以大脑为灵感的计算机芯片，可以通过更少的功率来提高人工智能的工作速度。他们的大型NorthPole（北极）处理器芯片不再需要频繁访问外部存储器，因此执行诸如图像识别之类的任务比现有架构更快，同时消耗的功率大大减少。10月19日，相关成果发表于《科学》。

它的能源效率令人震惊。法国巴黎萨克雷大学的纳米电子学研究员DamienQuerlioz说，这项研究表明，计算和内存可以大规模集成。我觉得这篇论文将动摇计算机体系结构中的普遍思维。

NorthPole运行神经网络：简单计算单元的多层阵列通过编程识别数据中的模式。底层接收数据，例如图像中的像素；每个连续层检测日益复杂的模式，并将信息传递至下一层。顶层产生输出，例如，可以表示图像包含猫、汽车或其他物体的可能性。

一些计算机芯片可以有效处理这些计算，但它们每次计算一层时仍然需要使用被称为RAM的外部存储器。以这种方式在芯片之间传输数据会减慢速度，这种现象被称为冯·诺依曼瓶颈，数学家约翰·冯·诺依曼是第一个提出基于处理单元和独立存储单元的计算机标准架构的人。

冯·诺依曼瓶颈是影响计算机应用（包括人工智能）发展的最重要因素之一，它还导致能源效率低下。论文作者之一、IBM的计算机工程师DharmendraModha说，他曾经估计，在这种架构上模拟人类大脑，可能需要相当于12个核反应堆的输出。

NorthPole由256个计算单元或核心组成，每个计算单元都有自己的内存。我们正在核心内克服冯·诺依曼瓶颈。Modha说。

Modha说，这些核心以一种网络连接在一起，这种网络的灵感来自于人类大脑皮层之间的白质连接。这一设计原则和其他设计原则——其中大部分以前就存在但从未结合在一个芯片中，使NorthPole在图像识别的标准基准测试中以相当大的优势击败了现有的人工智能机器。尽管它没有使用最新、最小型化的制造工艺，但它的能耗只有最先进的人工智能芯片的1/5。作者估计，如果北极设计采用最先进的制造工艺，其效率将是目前设计的25倍。

但是，即使是NorthPole有224兆字节的内存，也不足以容纳大型语言模型，比如聊天机器人Chat GPT使用的语言模型，即使在最精简的版本中，它也会占用数千兆字节的数据。

此外，该芯片只能运行预先编程的神经网络，这些神经网络需要在单独的机器上提前训练。但该论文的作者表示，NorthPole架构可能在速度很关键的应用中很有用，比如自动驾驶汽车。

---

NorthPole使存储单元在物理上尽可能接近核心中的计算元件。在其他地方，研究人员一直在使用新材料和制造工艺进行更激进的创新。这使得存储单元本身能够执行计算，原则上可以进一步提高速度和效率。（来源：中国科学报 文乐乐）

相关论文信息：<https://doi.org/10.1126/science.adh1174>

作者：Dharmendra Modha 来源：《科学》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](http://www.iikx.com)转发