

人工智能 | CPU 的性能超越 GPU，获京都大学青睐

原创：IDZ英特尔开发人员专区5月4日

京都大学医学院研究生院在进行计算药物发现时，利用 Theano 框架来训练深度学习神经网络，他们认为两路英特尔® 至强® E5-2699v3（Haswell 架构）系统的性能高于 NVIDIA K40 GPU 的性能。Theano 是一种 Python* 库，支持研究人员在 CPU 和 GPU 上透明地运行深度学习模型。其原理是使用针对目标架构的 Python* 脚本生成 C++ 代码。生成的 C++ 代码可以调用优化的数学函数库。

京都大学的团队认识到，开源 Theano C++ 多核代码的性能可以显著提升。他们和英特尔合作，共同提升多核性能。他们使用的是基于英特尔® 至强® 处理器的两路系统，因为当时新一代英特尔® 至强融核™ 处理器还没有发布。性能最终得以显著优化，证明在深度学习训练任务 1 中，两路 Haswell 处理器芯片的性能比 NVIDIA K40 GPU 的性能更出色。

在日本生物信息学协会 2015 年年度大会 (JSBI 2015) 上，Masatoshi Hamanaka（研究人员）展示了英特尔® 至强® 处理器的性能指标评测结果。基于这个结果和众核、多核运行环境的一致性，京都大学放弃 GPU，原因是 GPU 会增加不必要的成本、复杂性和内存限制，没有体现深度学习的性能优势。

演示稿中的“总结”页如下所示。



图 1: 与 GPU 相关的 Theano 优化加速和扩大英特尔至强内存容量的影响。

(结果由京都大学提供)

京都大学深度学习集群的采购是一个风向标，因为它是第一个选择多核 CPU 技术(而非 GPU 技术)的知名系统。期待 Theano 软件在下一代英特尔® 至强融核™ 处理器上的运行速度将显著提升。



图 2: 在 ISC'16, 英特尔详细展示了英特尔至强融核处理器在深度学习领域达到的卓越性能(与 GPU 对比)。

科学的重要性

京都大学医学院研究生院将各种机器学习和深度学习的算法应用于生命科学，包括药物发现、医学和医疗。 在其它领域，京都大学的研究人员需要处理大量的数据。 例如，京都大学团队希望将机器学习应用于实验技术产生的数据上，如高通量筛选 (HTS) 和新一代测序 (NGS)。 此外，可以用于分析日常临床实践产生的电子病例 (EHR)。 京都大学团队相信，通过使用大数据机器学习技术，相比之前的方法，他们能够执行更全面的分析。

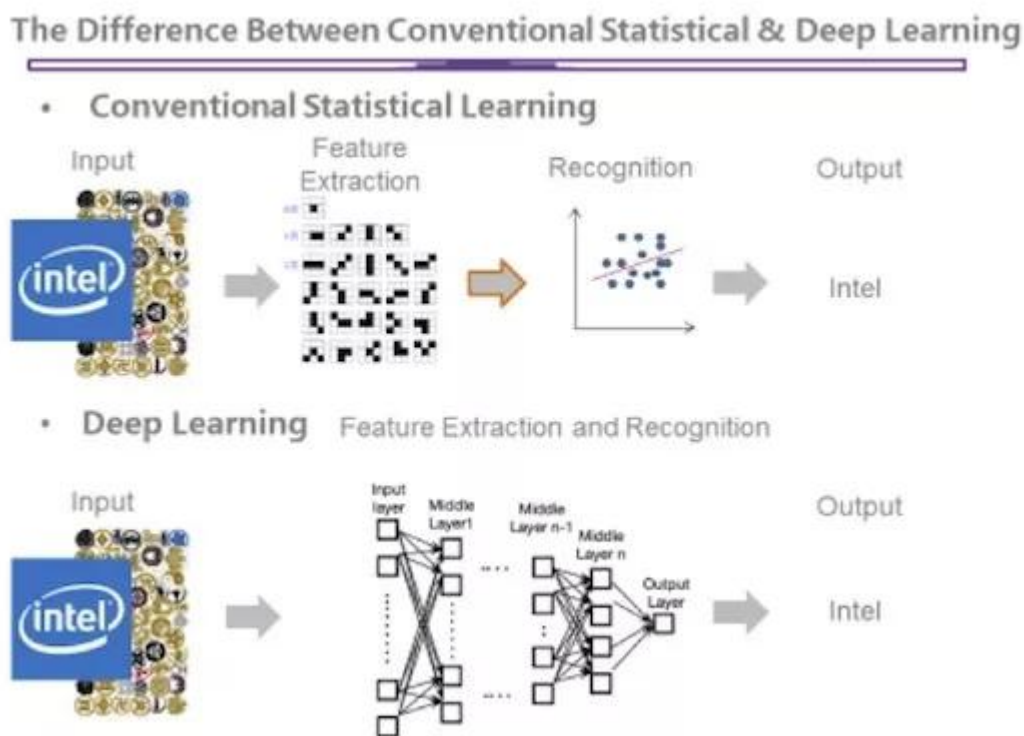


图 3: 图例显示深度学习和传统方法的区别。 (图像由京都大学提供)

在机器学习和深度学习研究方面，京都大学制订了 2 个目标：(1) 从迅速增长的数据中获得知识发现，数据包括实验数据和在病床边采集的电子数据。(2) 将知识发现的相关信息返回实验师和医生，完善药物发现和病人的医疗。

Yasushi Okuno 教授说：“未来 10 年，许多临床应用将会采用机器学习技术。”

“我们的机器学习和深度学习应用将在未来 10 年变得越来越重要。”

京都大学药物发现工作负载

京都大学工作负载的某些部分将计算虚拟筛选应用于药物发现领域。虚拟筛选在药物发现（耗时 10 年或更久）的早期阶段使用。虚拟筛选的目的是通过计算筛选大量的化合物，发现新的候选药物。

“目前，药物发现的早期阶段耗费几年时间，花费几亿美元。”Okuno 教授说。

“但是我们相信我们的研究将极大地节省时间和成本。”

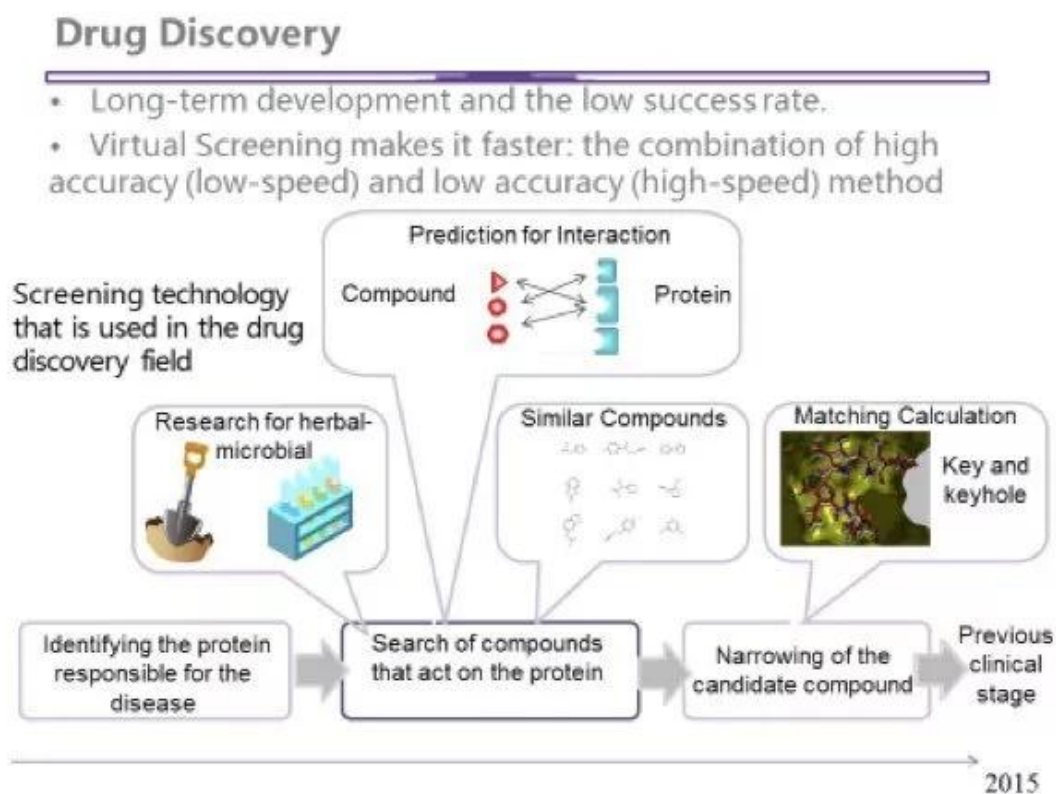


图 4：虚拟药物发现的案例与速度和容量有关。（图像由京都大学提供）

Okuno 教授说：“由于 DBN 根据数据学习，它可能会发现和现有的类药化合物结构不同的候选药物。” “因此，我们认为深度学习能够发现全新候选药物。”

Deep Learning (Deep Belief Networks)

- Learning the neural network of the deeper layers by pre-training
- The unsupervised learning with respect to the n-stage RBM (Restricted Boltzmann Machine)
- The final stage consists of a logistic regression
- Fine tuning the entire network (back propagation)

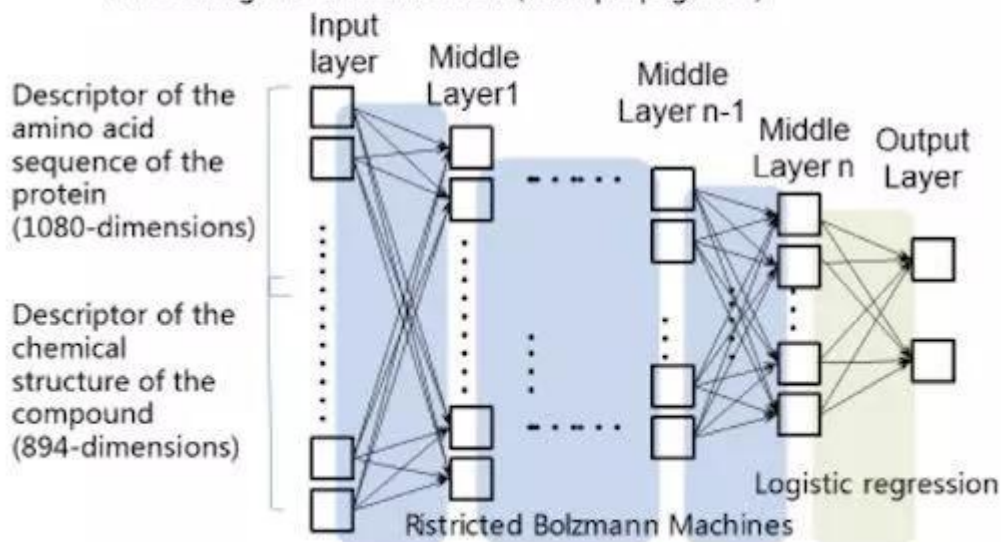


图 5: DBN 可以“学习”类药活动必需的数据特性。这些 DBN 能够用于预测、“评价”候选药物。（图像由京都大学提供）

大数据是准确训练神经网络进而解决复杂问题的关键。（在《神经网络的工作原理》中，Lapedes 和 Farber 认为神经网络实际上是一个‘坎坷不平’的多维平面，训练数据需要指定平面上的丘陵和山谷，或拐点。这就是为什么需要更多的数据以融入复杂的平面。）

Proposal of CGBVS-DBN method

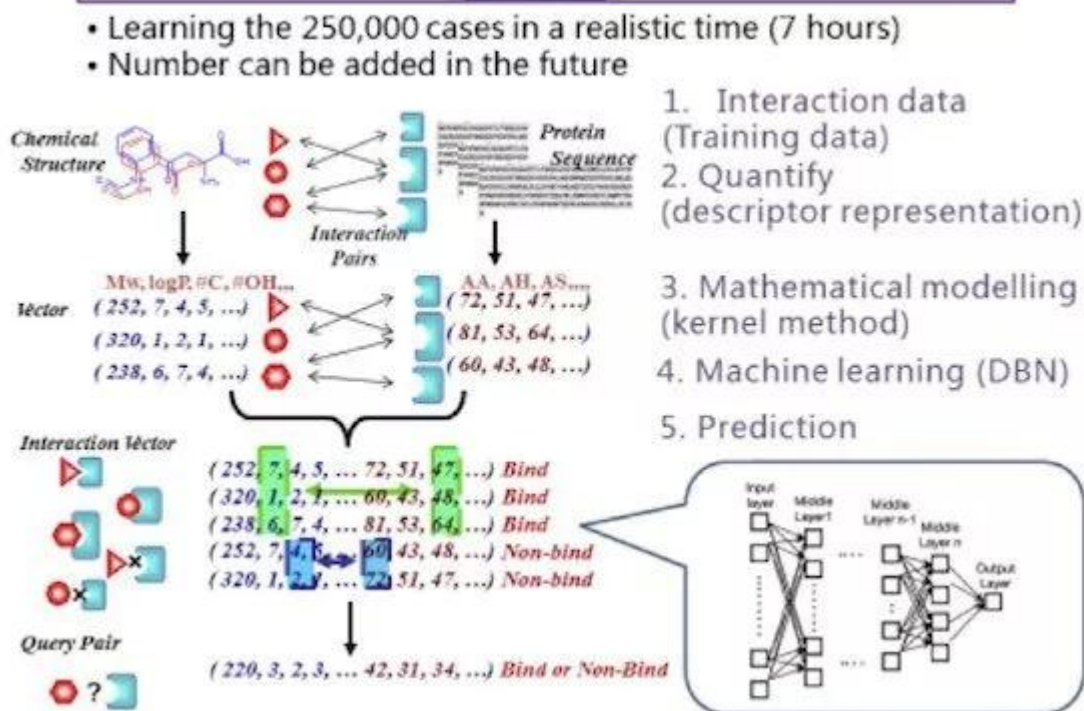


图 6: 利用深度学习来研发候选药物的推荐方法 (图像由京都大学提供)

京都大学数据集对 Theano 的扩展性能进行评估, 预计有 400 万行和 2000 个特性。利用 20% 的留存验证对结果进行验证。京都大学团队计划使用 Theano, 在包含 2 亿行、38 万个特征的数据集上训练, 数据增加了 130 倍!

“实验结果每天都在增加。”Okuno 教授说。“我们一直希望提高它们的计算性能。”

如下图所示, 随着数据集的扩大, 优化的多核 Theano 代码展现了出色的扩展性, 能够训练更多的数据。预计基于英特尔® 至强融核™ 处理器的新一代系统会进行类似的扩展, 提升性能和访问模型的速度。

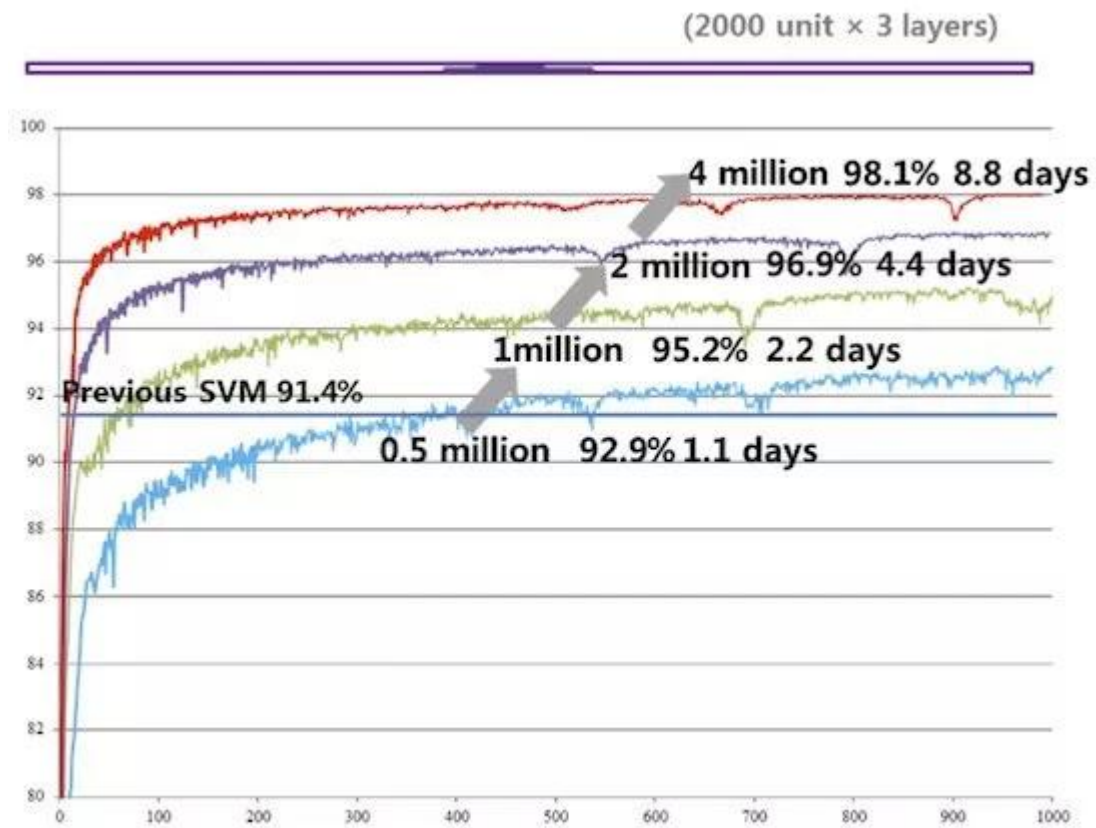


图 7：不同的数据规模下，优化的 DBN Theano 的扩展性。（图像由京都大学提供）

与 GPU 的代码路径相比，修复优化不佳的多核代码。

京都大学的结果显示，现代的多核处理技术与 GPU 机器学习性能相匹配，甚至超越了后者。但是同等优化程度的软件需要进行公平的性能指标评测对比。由于历史原因，许多软件包（如 Theano）缺乏优化的多核代码，这是因为所有的开源工作都投入到优化 GPU 代码路径中。

为了帮助他人进行公平的性能指标评测，为了实现多核、众核性能的优势，英特尔在 ISC'16 上，面向深度学习和机器学习推出了几个优化的库，如高级英特尔® 数据分析加速库（英特尔® DAAL）和针对神经网络（英特尔®

MKL-DNN) 库 (提供优化的深度学习基元) 的低级英特尔® 数学核心函数库。ISC'16 上发布的 MKL-DNN 库是开源的, 没有任何限制, 无需支付任何版税。成熟的英特尔® 数学核心函数库 (英特尔® MKL) 也添加了优化的基元, 更新机器学习, 以提升基于英特尔® 架构的机器学习和深度学习的速度。

例如, SGEMM 操作对机器学习算法至关重要, 后者大量使用单精度算法。新的库对 SGEMM 并行性进行了改善。

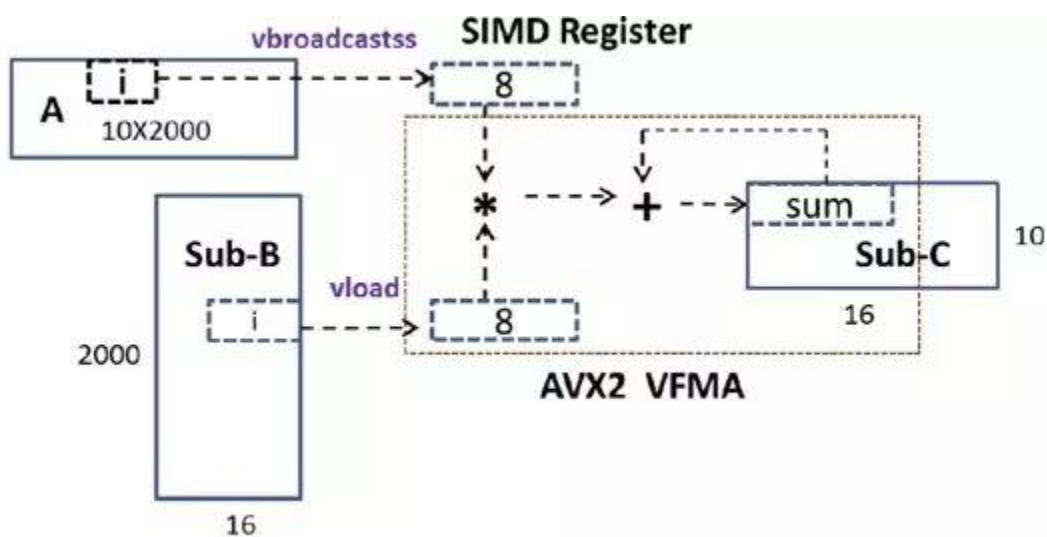


图 8: 利用 AVX 指令在相连地址对 SGEMM 并行性进行改善。(图像由英特尔提供)

ISC'16 上发布的针对矢量和多核优化的库将加快机器学习进度, 人们在评估机器学习的硬件平台时, 可以效仿京都大学, 利用针对多核优化的代码进行公平的对比。

基于英特尔至强融核处理器的新一代集群的预期目标

京都大学的计算和媒体研究学术中心 (ACCMS) 将采用基于英特尔至强® 融核™ 处理器的新一代集群，支持在更大规模的数据集上训练。具体的预期目标是：

1.

达到更高性能，和其他 CPU 和 GPU 平台相比，团队能花费更少的时间训练更大规模的数据。

2.

3.

协助发展高级算法。许多深度学习算法很复杂，这促使京都大学团队尽力消除架构的限制。多核扩展至众核的编程环境需要保持一致性，以消除 GPU 环境中的复杂性、内存限制和硬件偏差。此外，英特尔更新优化库的速度非常快，改善英特尔® 至强® 和英特尔® 至强融核™ 处理器性能。

4.

向人们普及多核处理器相对于 GPU 的优越性

为了帮助数据科学家和 HPC 社区了解并使用多核、众核软件和硬件技术，英特尔创建了机器学习门户网站，请点击 <http://intel.com/machinelearning>。网站将向读者介绍多核、众核处理器相对于 GPU 的优势，如优秀的训练和预测（也称作推断或评分）性能，以及在不同机器学习框架下出色的扩展性。通过这个网站，英特尔将为 100,000 名开发人员提供训练，使他们掌握机器学习技术和优化的库。顶级研究人员可以较早地获得技术信息，提供支持。