

# 脑科学视角下的高性能计算

刘亚东<sup>1,2)</sup> 胡德文<sup>1)</sup>

<sup>1)</sup>(国防科学技术大学机电工程与自动化学院 长沙 410073)

<sup>2)</sup>(国防科学技术大学国家高性能计算重点实验室 长沙 410073)

**摘要** 高性能计算已经成为脑研究的必要手段,为脑研究开启了全新的研究视角和研究模式;反过来脑研究为高性能计算提供新思路、提出了新要求,催生着新型计算模式的出现.具体来说:(1)近年来脑观测技术性能不断提升,多模态观测普遍采用,全球范围内交叉联合研究也越来越成为常态,使得“大数据”和“云计算”成为脑研究的一种崭新研究视角和研究方法;(2)信息科学家提出了深度神经网络、脉冲神经网络等大规模学习算法,这些算法从神经科学中汲取了营养,显著提升了自身性能,实现了人脑智能的某些特征,具有大脑计算模式高效率、低功耗的特征,在多领域内取得了显著的成效,这些算法的研究和实现对高性能计算提出了越来越高的要求,给目前基于传统微处理器和计算架构的计算平台带来了几乎无法克服的困难;(3)信息科学家和神经科学家联合,开始在类脑微处理器和类脑计算架构等层面上开展新型类脑计算探索性研究,取得的研究进展显示了该领域未来充满希望.基于以上三点原因,高性能计算和脑科学研究正越来越紧密地交织在一起,彼此融入了对方的发展.文中以脑研究为切入点,分别从高性能计算对脑研究的技术支撑以及脑研究对高性能计算的启发和推动两个方面对脑科学视角下的高性能计算的现状和发展趋势进行了讨论.

**关键词** 脑科学;高性能计算;大数据;类脑计算

**中图法分类号** TP18 **DOI号** 10.11897/SP.J.1016.2017.02148

## High Performance Computing in the Viewpoint of Brain Research

LIU Ya-Dong<sup>1,2)</sup> HU De-Wen<sup>1)</sup>

<sup>1)</sup>(College of Mechatronics and Automation, National University of Defense Technology, Changsha 410073)

<sup>2)</sup>(State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha 410073)

**Abstract** High performance computing has become a vital means of brain research, which opens up new research perspectives and patterns for it; In turn brain research gives rise to the emergence of a modern computing pattern by putting forward new requirements for high performance computing. Specifically: (1) in recent years, the performance of brain observation technology has steadily increased, the multimodal observation has been widely used, and the joint research worldwide has increasingly become the norm, thus “big data” and “cloud computing” become has been a new research perspective and methods of brain research; (2) large-scale learning algorithms such as large depth of deep neural network, spike neuron network, etc. were proposed by information scientists. These methods have learned much form neuroscience and improved their power significantly. The algorithms have high efficiency and lower power consumption characteristics of the brain computing model and implemented some of the characteristics of the human brain intelligence. And remarkable achievements have been made in many fields. The research and implementation of these algorithms have put forward higher requirements for high performance computing, which has brought almost insurmountable difficulties to the current computing platform based on the

traditional microprocessor and computing architecture; (3) information scientists and neuroscientists have jointly carried out a new type of exploratory studies at the level of the microprocessor and brain computing architectures, and the progress has shown a promising future. Because of the above three reasons, high performance computing and brain science are closely interweaving together and being integrated into the development of each other. With the background of brain research, this paper discussed present status and development trend of both high performance computing technology support for brain research as well as the brain research's inspire and promotion to the high performance computing.

**Keywords** brain science; high performance computing; big data; brain-inspired computing

## 1 引言

全球范围内脑科学研究方兴未艾. 欧盟<sup>①</sup>、美国<sup>②</sup>、日本<sup>[1]</sup>等均已制定了各自的脑研究计划, 中国也在脑科学上加强了部署. 国家“九七三”项目先后启动了“脑结构与功能的可塑性研究”、“人类智力的神经基础”等课题, 国家自然科学基金委启动了“视听觉信息的认知计算”、“情感和记忆的神经环路基础”等重大研究计划. 新技术、新手段、新发现不断涌现.

研究手段的创新与新技术的研发, 客观上为解开人脑之谜创造了条件, 其中高性能计算就是一项被寄予厚望的支撑性技术. 脑科学和高性能计算的关系是双向的: 高性能计算为迅速积累的脑观测数据提供了处理方法与研究平台, 推动脑研究进入了“多模态、大样本和联合研究”的时代, 并促使新的发现和新的数据不断涌现. 海量数据蕴含着更多的信息价值, 为后续的脑科学研究提供了重要支撑. 反过来当前的脑科学研究不断对高性能计算技术提出全新的性能要求, 成为推动高性能计算平台计算效能提升的动力之一. 此外脑科学为信息科学提供了新的方法论, 催生着对深度神经网络、脉冲神经网络、类脑计算芯片等新型高性能计算方法和平台的研究, 不断为“类脑计算(Brain-inspired computing)”这一研究领域提供营养.

脑科学离不开高性能计算的技术支持. 不论是利用高性能计算平台完成大规模数据处理工作, 还是将其运用在大脑的计算与建模, 高性能计算技术都极大地加速了脑研究的进程. 尤其在基于信息通讯技术的新型脑研究模式下, 科学家可以利用日益先进的高性能计算技术构建智能信息处理平台, 凝练、聚合脑科学相关研究成果, 对大脑进行计算模拟, 从而在多尺度上探寻大脑的工作机制、正确认识脑疾病病理, 并在高性能计算平台上实现强人工智能.

本文探讨了高性能计算与脑科学目前交织发展的现状和未来的发展趋势. 本文第 2 节指出脑科学研究已步入“多模态、大样本和联合研究”时代, 对催生海量数据的各种脑科学研究计划以及与之相关的大数据等高性能计算发展现状进行了梳理和分析; 第 3 节对深度神经网络、脉冲神经网络、皮层学习算法、类脑计算芯片等发展领域进行了讨论, 特别分析了这些类脑计算方法对皮层神经网络的借鉴; 第 4 节讨论类脑计算的意义、存在的问题以及未来可能的发展方向; 第 5 节对全文进行总结. 图 1 给出了高性能计算和脑信息处理机制研究的关系, 其中涉及的具体技术在论文的适当位置都有论述.

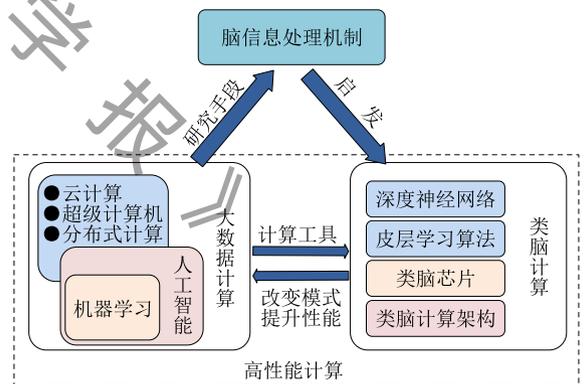


图 1 高性能计算与脑信息处理机制研究的关系

## 2 高性能计算已经成为脑科学研究的必要手段

### 2.1 脑科学研究步入“多模态、大样本和联合研究”时代

全球脑科学研究已经具有了相当的规模, 每时

① The Human Brain Project Begins. <http://www.forbes.com/sites/jenn-iferhicks/2013/10/07/the-human-brain-project-begins/>  
 ② BRAIN initiative. <http://www.whitehouse.gov/share/brain-initiative>

每刻都在产生着大量的数据.但是数据的飞速增长并不一定带来知识的飞速增长.不同的数据只是从不同的侧面观测大脑的结果,这样的观测视角是狭窄的.就好像我们的目的是研究白光的光谱分布,但不同的人看到的只是被三棱镜分光的特定波段的光谱,这样一来,研究再细致,我们也无法回答白色光谱的分布问题.作答的唯一途径就是综合不同波段下的观测,才能形成对白光的统一认识.目前的脑科学研究也是这样的,我们在每个层次上细致研究,但缺乏对所有研究的综合.我们必须集合可以利用的数据,建立一致、统一的框架来处理、解释这些数据,才可能最终回答大脑智能信息处理的奥秘.综合已有的和未来的海量数据,从中寻找规律,提出框架,我们就需要高性能计算技术的支持.原因有:

(1) 脑观测工具性能的不不断提升正产生着更多的数据<sup>①</sup>:随着脑研究手段的多样化和观测工具性能(特别是时空分辨率)的逐步提升,科学家面对着快速增加的海量数据.例如,fMRI 采集周期从 20 世纪 90 年代的 4 秒已经变成了目前的不足 1 秒;空间分辨率也从最初的 5 mm 立方进展到了目前的 0.5 mm 立方.此外在分析解释新数据时科学家往往需要从过去的数据中获得先验知识,这也使得科学家面临的数据量进一步增加<sup>[2]</sup>,大数据已经成为脑科学研究绕不开的问题.例如,2013 年 6 月,《科学》杂志介绍了德国 Juelich 研究中心的 Amunts 和她的团队完成的名为“BigBrain”的超高分辨率 3D 人脑模型<sup>[3]</sup>.她们将一名 65 岁死者尸体的大脑分割成了 7400 层 20 微米厚的切片,之后对切片进行染色,并使用标准实验室相机对每个切片拍摄高分辨率(13000×11000 像素)图像,得到一个 1 万亿字节的全脑体元数据集,再利用 Compute Canada 网络提供的高性能计算设备实现了数据集的可视化,最终得到一个超高分辨率的 3D 人脑模型.“BigBrain”也可供公众免费使用<sup>[3]</sup>,研究者们可以利用该模型进行数据仿真与实验假设验证.

(2) 多中心、多时段数据分析的需求.不同研究中心积累了很多的历史数据,单个来看,每个数据集规模都不大,但如果能够集中起来加以利用,我们将得到很多的大数据库.利用这些数据库,从大样本角度,我们就可以开展在单中心无法完成的研究工作.所以这些分散的数据是一笔潜在的财富.目前的难点并不仅仅在于将多中心、多时段的数据收集起来,现有分析方法是针对少量被试、单一中心、单一时间

段数据设计的,不适用于多中心、多时段数据的处理,急需信息科学家提出新的方法实现多中心、多时段数据处理.研究发现,即使相同品牌的采集设备,不同的实验室产生的数据质量也存在差别<sup>[2]</sup>.此外很多历史研究关注的是脑结构和功能随时间的变化,这一时间跨度可能长达几年,甚至一生,这也会造成全部数据来自于不同的设备、不同的时段的结果.

(3) 多种观测手段的联合应用.多种脑观测手段的联合采用已经成为脑研究的一种方式.2010 年,美国启动了“人类连接组计划”(Human Connectome Project, HCP)<sup>②</sup>,目前向公众开放了近 50% 数据(500 余名被试),每个被试的数据集包括静息态功能磁共振数据、弥散磁共振数据、任务相关功能磁共振数据、脑结构磁共振数据 4 种,部分被试还将提供脑磁图数据<sup>②</sup>,目前共有 20T 的数据可以在网上访问.基因和脑成像这两种以前分列的研究方法目前也实现了联合.例如,2011 年 4 月,美国艾伦脑科学研究所(Allen Institute for Brain Science)成功绘制出了两个迄今最完整的人脑基因图谱,并将相关研究数据编制成一个名为“艾伦人脑图谱”(Allen Human Brain Atlas)的公共数据库<sup>[4]</sup>.该图谱不仅显著标识出了人类特定基因与大脑特定结构和区域的映射关系,还涵盖了大脑核磁共振成像与弥散张量成像的数据<sup>[1]</sup>.基因成像和人脑成像技术的飞速发展,产生了越来越多的数据,两者的结合就意味着大数据.在多模态研究中,无论是数据量存储本身,还是后续的传输、压缩、计算等,都给信息分析带来了巨大的压力.对一些基础性脑疾病,如帕金森、精神分裂症等,多角度的研究已经开始<sup>[5]</sup>.这些都需要多研究中心的数据配合,并且需要制定一些整合和揭示数据的规则,以协调不同的研究<sup>[6]</sup>.

(4) 全球范围内不断开展的各类脑研究计划:不论是以静态大脑解剖结构为研究内容,以获得全脑映射图谱的相关脑研究计划,还是着眼于记录神经元的动态活动,以探寻大脑功能和行为之间复杂联系的脑研究计划,都在为脑科学研究积累数据.

美国 HCP 计划使用不同的脑成像技术,通过对 1200 名健康成年人(包括 300 对双胞胎)的大脑扫描数据进行辨别与分析,旨在绘制出人类大脑生

① New ways to image neural activity. <http://grants.nih.gov/grants/guid-e/rfa-files/RFA-EB-05-001.html>

② The human connectome project. <http://www.humanconnectome.org/>

物线路的分布与连接情况<sup>①[7]</sup>. 2013 年启动的欧盟“人脑工程”计划(the Human Brain Project, HBP)<sup>②③</sup>, 拟对神经科学数据进行整合, 在神经元的层次上模拟脑的行为, 并希望最终建立起一个供神经科学家进行研究的模型工具<sup>④</sup>. 该计划目前设立了 13 个子项目, 其中第五个子项目“神经信息学”就是为联合研究建立数据收集、共享与分析平台<sup>⑤</sup>. 同时建立多层级脑图谱以及脑结构和功能数据收集与分析新型工具. 数据将来自于往过文献和实验. 数据将被仔细描述以方便实用. 项目还将发展大数据分析手段, 从海量数据中挖掘数据模式和生物组织行为的关系. 数据还将涵盖各类离子通道、各类分子、各类突触以及神经微回路的观测数据, 如电子显微镜数据、局部场电位数据、弥散张量数据等<sup>⑥</sup>. HBP 同样以整合已有神经科学数据和知识, 在高性能计算机上模拟人脑为目标. 2015 年, HBP 研究团队结合多种神经元观测手段, 利用高性能计算技术实现了对大鼠体感运动皮层神经元网络初步的数字化模拟(First-draft). 模拟建立在对皮层神经元形态、功能、电特性、生化特性等细致观测的基础上, 得到的模拟神经网络具备某些真实神经网络的动力学特性. 此次模拟使用的高性能计算平台包含 4 台 IBM 蓝色基因 Q 型超级计算机, 同时使用超过 4.2 皮字节的存储空间存储原始数据<sup>⑦</sup>.

多模态观测、大样本数据收集以及大范围联合研究使得高性能计算成为研究的必要手段. 庞大数据量带来了数据存储与分析方面的难题, 需要“大数据技术”作为支撑. 在基于信息通讯技术的新型脑研究模式下, 如何有效地将脑科学研究与高性能计算技术相结合是目前亟待解决的问题.

## 2.2 脑研究大数据的存储与分析

脑研究产生的大数据为科学家的深入研究提供了海量信息. 然而, 大数据的存储与分析却面临着诸多困难, 如海量数据的安全廉价存储、快速访问、高效处理等, 这些都要求研究者们探寻新的数据处理模式<sup>[2]</sup>. 显然目前快速发展的“大数据”、“云计算”等技术可为当前脑科学的研究提供技术支撑.

大数据技术的目的是进行海量数据的收集、存储和处理, 并在此基础上进行规律判断和趋势预测. 大数据的特点可总结为: Volume(容量巨大)、Variety(类型多样)、Value(信息价值高)、Velocity(传输与处理速度快). 大数据时代, 从方法论角度来看, “相关性分析”已经成为一种重要的数据分析方

式. 如果面对科学问题, 特别是复杂科学问题时, 直接的因果关系不可获得, 研究者可以退而求其次, 通过数据的模式分析, 建立模式间稳定的“相关性”, 利用这种相关性, 研究者可以启发后续的研究路径, 或者对未来趋势做出预测. 也就是说, 如果彻底的理解不可得, 那么我们就去关注某些现象在历史数据中以什么样的规律反应出来. 如果这个规律反复稳定的出现, 我们就将它作为事实接受下来. 大数据下数据具有了绝对重要的地位, 当数据多到几乎可以均匀覆盖整个样本空间, 那么数据就可以被看做是对“真实情况”的采样, 只要采样足够细密, 基于数据的判断和预测都变成了一个“内插”的过程, 并且内插可以带来足够的精度.

学术界密切关注大数据相关问题, 各类顶级期刊纷纷设置大数据专刊, 如《Nature》于 2008 年推出“Big Data”专刊<sup>[9]</sup>, 《Science》则在 2011 年 2 月推出“Dealing with Data”专刊<sup>[10]</sup>, 对大数据带来的相应技术需求及所面临的挑战进行探讨. 其中, 2013 年 6 月《Nature》的“Big Data”专刊中名为“大数据面临大挑战”的文章<sup>[2]</sup>, 针对大数据背景下生物学家和计算机科学家进行大数据存储、对比、分析与共享的方式, 以及处理大数据所遇到的瓶颈等问题进行了剖析. 文章指出, 针对大数据生物学问题, 依靠网络以及云, 而非本地软、硬件去进行数据的存储与处理是必然趋势. 这里, “云计算”的概念被提及, 已然成为解决如人类基因组计划、脑科学研究等大数据生物学问题的关键技术.

云计算是大数据的必然结果, 它的出现减少了科学工作者在计算、数据、实验条件等方面的研究障碍, 一流实验室的软硬件研究条件得以共享. 利用云计算进行数据存储、处理与分析有如下几点优势: (1) 运用数据并行处理方式, 将巨大的运算任务细分给云端中的大量机器同时进行运算, 大幅提高了大数据的分析速度, 有效解决了大数据处理运算耗时的问题<sup>[10]</sup>; (2) 研究人员不用再重复消耗大量资源、建立独立的软硬件设施和维护人员队伍, 而可以按需购买云计算服务, 减少了数据存储与处理成

① The human connectome project. <http://www.humanconnectome.org/>

② The Human Brain Project Begins. <http://www.forbes.com/sites/jenn-iferhicks/2013/10/07/the-human-brain-project-begins/>

③ Human brain project — A report to the European Commission. <http://www.humanbrainproject.eu>

④ SP5-neuroinformatics. <https://www.humanbrainproject.eu/neuroinformatics-platform>

本及人力成本；(3)降低计算和数据获取的难度，将研究力量集中在研究方法、研究思维上，避免研究的低水平重复<sup>[2]</sup>，有利于促进脑科学研究的全球化合作。大数据技术可使研究者重新检视过往实验研究数据，这包括自己的和他人的研究。这样检视的结果往往是新结论、新发现的出现。并且当许多人面对着相同的材料进行思考时，可能会产生出类似于头脑风暴的东西来。这无疑是对研究的有效推进，对于单个研究者来说显然也是有益处的。

当前，全球已经有多个针对脑科学研究的云服务平台开始建设，使得脑科学研究依托云计算平台进行数据的存储与分析变得更为便捷。欧盟 HBP 计划的第五个子项目“神经信息学”，就计划建立多模态的脑数据存储和分析的云平台<sup>①</sup>。平台收集历史研究数据以及未来的实验数据，并加以统一描述。平台还提供数据收集、解释和分析的工具，以方便访问者的使用。近期为研究脑小尺度和长程连接的弥散张量数据库也将要完成。

2013 年奥巴马宣布启动“BRAIN—Brain Research through Advancing Innovative Neurotechnologies”研究计划<sup>②[11]</sup>，其中的 U24 子项目称为“The Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC)，神经成像信息学工具与资源交换”<sup>③</sup>，目的就是“为研究者和开发者提供访问、信息、论坛等平台，以实现数据等资源和数据分析等工具的“增强、共享、应用、演进”。项目分为三个部分：NITRC-R 部分提供软件和工具交换；NITRC-IR 提供神经成像数据的公共访问；NITRC-CE 部分为全球研究者提供针对神经成像数据库的云计算服务<sup>④</sup>。美国 HCP 计划也开放了数据库的访问，提供了针对磁共振数据和脑磁图数据的免费分析工具<sup>⑤</sup>。

我国在“十二五”规划纲要中明确指出要大力发展云计算。由科技部批准的、已建成或正在建设的云计算中心有北京、无锡、深圳、长沙、天津、济南、青岛七处。在全球范围内，云计算领域已出现了合作趋势，一些主要的计算中心和机构建立起了一定程度的联合机制，如加拿大七个计算中心整合的 Compute Canada 项目，欧洲的 PRACE 项目等<sup>[12]</sup>；这些项目的开展将进一步加强云计算服务平台的资源共享能力，并方便进行联合调度。

脑科学研究的高度复杂性需要全球性合作，而云计算技术可以加强脑科学研究数据的共享能力，有利于促进具有一定异构性的科学研究与临床数据向集中和统一的方向发展。云计算和存储本身也带

来了新的技术难题。如数据的安全性，对平台的信任程度等。数据传输速度也成为一个需要考虑的难题<sup>[2]</sup>。云计算的免费化是目前的发展趋势之一，并且这种计算服务应该减少用户对专业知识的依赖，进而让大数据技术为多领域提供新的技术碰撞的机会<sup>[2]</sup>。

### 2.3 计算能力在脑科学研究中的作用日益凸显

如今，科学研究已离不开计算机技术的支持；尤其在大数据背景下，面对蕴含丰富信息的海量数据，高性能计算机已经成为大规模数据处理与高精度仿真的必要工具。

脑科学研究与高性能计算技术的联系更为紧密。研究者们主要利用高性能计算技术来完成对数据处理分析、模型建立和求解等工作。例如，在绘制数字化 3D 人脑图谱的过程中，科学家要利用高性能计算机的“Brute Force Rendering，蛮力绘制法”完成数据集的可视化工作<sup>[13]</sup>；在进行弥散张量成像数据的分析时，国外研究机构多使用集群计算技术来实现。国内基于弥散张量成像的脑研究，也开始使用高性能计算平台。如北京师范大学认知神经科学与学习国家重点实验室与北京师范大学计算中心合作，利用计算中心的高性能计算机进行弥散张量成像纤维跟踪的计算分析。

欧洲 HBP 计划目前设置了 4 台超级计算机为脑研究服务<sup>⑥</sup>。计划中的第七子项目“High Performance Computing(高性能计算)”，专门研究超级计算技术在脑科学研究中的应用。核心研究内容包括“新型加速技术”、“分层存储”、“百亿亿次交互计算”、“可视化技术”、“大数据整合”。这一子项目的目标之一是为 HBP 计划提供超算能力，特别是针对实现多尺度脑模拟模型的建立，在“大数据计算”、“云计算”、“交互计算”、“可视化”等方面提供技术支撑<sup>⑦</sup>。计划到 2017 年实现 50 皮字节分层存储、每秒 5 亿亿次浮点运算、总功耗小于 4 兆瓦的超级计算平台。这样的计算能力将能够模拟一只小鼠的大脑或者实现对人脑的初步模拟。到 2022 年实现 200 皮字节分层存储、每秒 1 百京( $10^{18}$ )次浮点运算的超级计算平台，具备对人脑的多层级的模拟能力<sup>⑧</sup>。

① SP5-neuroinformatics. <https://www.humanbrainproject.eu/neuroinformatics-platform>

② Interim report. <http://www.nih.gov/science/brain/>

③ Neuroimaging informatics tools and resources clearinghouse. <http://grants.nih.gov/>

④ The human connectome project. <http://www.humanconnectome.org/>

⑤ SP7-High Performance Platform. <https://www.humanbrainproject.e-u/high-performance-computing-platform>

然而,若想要充分利用大数据的价值与预测功能,把对大数据的加工处理转变为对大数据的最终理解,高性能计算机在研究中的作用范围就必须予以拓宽. 21 世纪科技发展的一大趋势,便是将计算机支持科学家做传统科学研究转向计算机嵌入到科学研究的全过程,形成“计算+传统科学=新科学”的新局面. 面对这一种新的科研形式,一些研究者将目光聚焦于脑研究与高性能计算技术的结合,提出了基于信息通讯技术(Information and Communications Technologies, ICT)的新型脑研究模式<sup>[14]</sup>. 在这种新型模式下,过于零散的脑研究数据将被整合,进而建立起详细的计算机模型,对人脑的活动进行模拟.

### 3 脑启发下的新型高性能计算

除了纯科学研究,研究大脑的主要目的是希望实现机器对人脑的模拟,使机器像人一样具备高度智能,具备高效率低功耗的信息处理能力. 人工智能和人工神经网络领域就是对人脑智能模拟的一种尝试. 人工智能是通过程序编制使计算机实现人脑的某些智能特征. 但方法中对智能的借鉴只是很少程度上来源于对人脑神经网络的知识. 更多的则是通过借鉴人类智能行为规则来设计规则,再让计算机利用这些规则进行推理以实现对类似情况的处理. 如医学领域的计算机辅助诊断系统,就是将医疗专家对不同情况的处理规则进行程序编制实现,使计算机能够进行辅助诊断和对治疗方法进行建议. 但是,将人脑的思维过程简单理解为一个基于规则的推理过程是不准确的,如在处理复杂或者紧急情况时,专家直觉处于一个相当重要的地位. 即使不考虑“直觉”这种难以抽象和建模的智能过程,推理过程规则的建立也是件相当困难的事情. 人工智能的能力是建立在基于冯·诺依曼计算平台的算法和程序之上的,虽然可以用来解决目标跟踪、可通行区域识别、棋类竞技、智能检索等复杂问题,可是和人脑的智能相比还处在一个相当低的发展水平. 其原因可能是人脑神经网络和计算机的集成晶体管在计算模式上存在着根本差别. 和人工智能不同,人工神经网络的研究者希望通过建立网络连接模拟的人脑神经网络,以实现某些人脑的智能和计算能力. 这些研究者认为网络节点间的连接关系及其连接权重就是智能实现的基础,他们的观点被称为“联结主义”. 和基于冯·诺依曼计算平台的人工智能不同,神经网络没有计算中心(CPU)和存储中心(硬

盘),计算和存储分散在节点和通路上. 人工神经网络的成功是非常有限的,只在手写字符识别等有限问题上取得效果. 表面看起来,人工神经网络看似是对脑的模拟,但这一模拟在规模和连接方式上即使跟相对简单的果蝇神经网络相比还有很大差距,更遑论人脑网络. 具体来说两者有如下区别: (1) 早期人工神经网络仅有数十或者数百神经元,人脑则有 860 亿个; (2) 人脑神经网络中存在着复杂的反馈连接,反馈连接是前馈连接数量的数倍至数十倍,这一点仅在 Hopfield 神经网络、模拟退火算法等部分人工神经网络中有所体现; (3) 人脑智能的飞跃和新皮层的出现关系密切,新皮层使人类具有了语言、逻辑、抽象思维等智能行为,而人工神经网络并没有针对性的对新皮层进行模拟; (4) 可能是最重要的一点,人工神经网络被设计用来处理静态信息,如图像处理等,而人脑处理的五种感觉信息基本上都是随时间变化的信息. 加入了时间维度,动态信息处理显然和静态信息处理的机制存在显著区别,所以从功能角度来看,两种网络也是有显著区别的. 直到近期,处理序列信息的深度神经网络才开始出现,如 Recurrent Neural Networks(RNN)开始涉及文本理解、文本翻译等序列信息处理<sup>[15]</sup>,但总体来看神经网络依然主要应用于静态信息的处理. 综上,总体看此人工神经网络非彼人脑神经网络,人工神经网络并未完全体现出人脑神经网络的精髓.

#### 3.1 类脑计算研究兴起

近十年来,模拟人脑神经网络以实现智能这一研究思路再次活跃,原因之一就是近 20 年来软硬件计算能力的飞速发展,使得大规模人工神经网络以及基于类脑芯片的新型计算平台的研究成为可能. 这些研究一方面在体系结构和底层功能上更多地借鉴了人脑,力图对人脑神经网络进行“逼真”模拟,另一方面在网络规模上也向人脑看齐. 由于结构的复杂性和网络规模的迅速提升,使得网络的建立、训练和应用都需要高性能计算平台作为工具. 另一方面这些研究可能突破冯·诺依曼的计算架构,为高性能计算平台带来了变革契机.

##### 3.1.1 模拟脑成为各脑研究计划的目标之一

如果要实现具有人脑智能水平的信息处理系统,一方面,要综合关于人脑神经网络结构和功能的已有知识,同时要继续推动相关研究,在多层面上展开对人脑神经网络生理和动力学的研究,以期不断获取新的知识,使人工网络更加“逼真”;另一方

面要利用高性能计算技术实现超大规模的网络建立、训练和应用. 分散的研究团队不具备开展这种规模研究的支撑条件和学术背景, 需要在国家层面上制定统一协调的研究计划.

欧盟 HBP 计划的一个主要目标就是利用从大脑得到的模型发展新型计算技术<sup>①</sup>. 计划的第六个子项目称为“Brain Simulation(脑模拟)”, 项目将在抽象计算模型、单点神经元模型、分子水平上的神经环路和小脑区模型以及可动态切换多层次描述模型等多个方面展开对大脑的模拟<sup>②③</sup>. HBP 拟在基因、递质、神经元、环路、脑区、全脑网络等层面上同时对大脑展开研究, 这些研究是对“脑模拟”项目的有力支撑, 获得的新知识将使模拟网络更加逼真. 除了研究大脑, HBP 的另一个重要目的是利用大脑, 研究类脑计算技术, 实现智能、高效、低功耗的计算技术. 计划 2022 年在计算平台中引入人类脑的信息交互和数据计算子系统, 实现百亿亿次超级计算<sup>④</sup>. 美国国防部高级研究计划局(DARPA)支持的“神经形态自适应可塑可扩展电子系统”项目, 在“微处理器”和“计算架构”两个层次上模拟人脑神经网络的信息处理机制<sup>⑤</sup>. 美国情报高级研究计划署是“基于皮层网络的机器智能”研究计划的资助者. 资助此研究计划目的是“逆向大脑的信息处理规则, 彻底改变机器学习”. “参与者将利用他们对大脑的知识表征、转换和学习规则的理解, 创造能力更强的类脑机器学习算法”<sup>⑥</sup>. 这一机构历来感兴趣的是那些即将投入使用的顶尖技术研发.

在国内, 2012 年复旦大学联合十几所高校及中科院研究所, 成立了“脑科学协同创新中心”. 2015 年清华大学成立了“类脑计算研究中心”, 目标是借鉴人脑信息处理方式, 实现具备非结构化信息实时处理和学习能力的超低功耗新型计算系统. 2015 年中国科学院自动化研究所成立了“类脑智能研究中心”, 其目标是研究创新性的认知脑模型, 实现类脑信息处理、类脑智能机器人等相关领域理论、方法与应用. 目前已建立小猫部分视皮层的方向选择神经网络模拟系统、小鼠的记忆模拟系统, 其中包括对不同脑区、不同类型神经元形态和工作模式的建模与模拟.

### 3.1.2 模拟脑不是对脑皮层的完全复现

要实现类脑计算, 显然模拟大脑的信息处理方式是一个直接并且有效的方式. 问题是我们如何借鉴? 在什么层次上去借鉴人类大脑? 完全复制大脑是不可能的, 也是没有必要的. 聪明的做法是搞明白

脑信息处理的一般原则, 并在计算中应用这些原则, 而实现的途径和技术路线完全是可以根据实际情况来做决定的.

脑信息处理有两个基本原则: 一个是“层次结构, Hierarchies”<sup>⑦</sup>, 另一个是“序列预测, Sequence Prediction”<sup>⑧⑨</sup>. 所谓“层次结构”是指人脑神经网络是分层构造的. 每层的结构并没有显著的区别. 人脑大规模的神经网络是通过一些简单的神经网络逐层叠加和空间扩展形成的. 不同认知功能的神经网络也具备着类似的分层结构. 如果将视觉信号输入听觉皮层, 听觉皮层将学习去“看”<sup>⑩</sup>, 如果把视觉信号输入运动感觉皮层, 运动感觉皮层也将学习去“看”. 可以想见的是不同感觉通路的皮层神经网络信息处理机制是相似的, 或者说他们从输入信号中提取信息的基本原则是相同的. 所以, 相同结构的神经网络是可以拿来处理不同领域的数据的<sup>⑪</sup>. 所谓的“序列预测”是指人脑神经网络具备“时间观念”, 能够从数据中提取时间上先后发生的模式. 正是因为能够学得这种先后模式, 建立所谓“因果链”的诸多经验, 才能从过去的生活中汲取知识来指导未来的生活. 正确使用这些简单原则就有望构造出大规模类脑计算神经网络. 例如“深度神经网络”(Deep Learning Neural Network, DLNN)计算方法借鉴了“层次结构”原则<sup>⑫</sup>, “皮层学习算法”(Cortical Learning Algorithm, CLA)<sup>⑬</sup>、脉冲神经网络(Spike Neuron Network, SNN)<sup>⑭</sup>等借鉴了“序列预测”原则.

大脑的优越性不在于规则的复杂性, 而在于其超大的网络规模. 目前的软硬件计算技术是可以模拟大规模网络的, 所以在使计算具有人类智能这件事上我们有越来越多的事情可以做<sup>⑮</sup>. 相对于人脑皮层神经网络的规模, 目前绝大部分人工神经网络规模要小的多.

在思考机器智能的时候, 我们可以将人类智能作为参照, 甚至作为最重要的参照, 但要意识到, 模拟智能的目的并不是完全实现人类认知行为能力的完全复制, 而是可以在诸多方面超越人类智能. 如在

① SP11-Applications. <https://www.humanbrainproject.eu/applications>

② Human brain project — A report to the European Commission. <http://www.humanbrainproject.eu>

③ SP6-Brain Simulation. <https://www.humanbrainproject.eu/brain-simulation-platform>

④ Neuroscience Programs at ARPA. <http://www.iarpa.gov/index.php/research-programs/neuroscience-programs-at-iarpa>

思考速度、记忆容量、感知能力、系统复杂度、可复制性等诸多方面实现对人类智能的超越。神经元是以  $1/10^3$  秒的速度工作, 而硅基芯片则以  $1/10^{10}$  速度在工作, 它们的差别是 100 万倍, 这意味着理论上如果芯片具备了思考能力, 它的思考速度是人类的 100 万倍<sup>[18]</sup>。这将给我们的生活带来很多不可思议的变化。不管人脑确切的记忆容量是多大, 智能机器的记忆容量都是可以远远的超过它的。所以类脑计算的目标不会停留在复现大脑智能计算能力, 而应该是实现按照大脑方式进行计算, 诸多方面性能又优于人脑的计算体系。

### 3.2 大规模类脑神经网络

人脑智能来自于其超大规模、复杂互联的神经网络, 模拟这个网络是实现类脑计算的主要思路之一。在人工智能等发展受挫以及大规模网络构造与计算具备可行性后, 这一思路受到越来越多研究者的重视。目前大规模神经网络有两个主要的发展思路, 第一个发展思路是在信息处理原则上借鉴人类脑皮层, 在规模上向人脑看齐, 以实现人脑智能, 如深度神经网络、皮层学习算法等。这类网络在具体细节上并不刻意模拟人脑神经网络, 他们依然归属于“人工神经网络”的范畴; 第二个发展思路是从单个神经元设计到网络结构都尽可能的模拟人脑皮层神经网络, 如 SNN 等。这类神经网络属于“脑启发的神经网络”, 希望通过“逼近”大脑实现人脑智能。

#### 3.2.1 深度神经网络的“层次结构”

深度神经网络是一种类脑计算方法。斯坦福大学的 Ng 教授认为和神经科学家合作并从神经科学研究中汲取营养的确提升了深度神经网络的性能<sup>[17]</sup>。深度神经网络和人脑皮层神经网络是有相似处的, 这种相似处即吸引着信息科学家, 也吸引着神经科学家。如深度神经网络中的“卷积神经网络(Convolutional Neural Networks, CNN)”, 其中的卷积层和汇聚层(Pooling layer)就是受到视觉皮层简单神经元和复杂神经元的启发, 而网络的整体结构和视觉通路 LGN-V1-V2-V4-IT 的结构类似<sup>[23-25]</sup>。总体来说, 人脑皮层和深度神经网络的相似之处有二, 第一个相似之处是: 他们都具有从新数据中学习知识的能力。人脑的大部分知识不是从父母和老师处学到的, 而是在外部世界中自己不断探索得到的。深度神经网络也是这样, 能够从新数据中获得新知。借鉴人类视觉系统多尺度特征提取的特点, 有学者提出了 HCNN 网络<sup>[26]</sup>, 将 CNN 和梯

度直方图特征结合, 在提取图像局部特征的同时, 提取出更多的全局性特征, 提升了 CNN 在目标识别中的应用效果。第二个相似之处是: 深度神经网络和人脑神经网络都是由简单的神经网络通过分层结构“累加”而成的。每层子网络完成的是从输入层到输出的映射, 并且这种映射在大部分情况下都是非线性的。随着层级的增加, 层子网络学习得到的特征针对分类这一任务来说“特征选择性”和“特征抽象性”都将逐渐加强。如深度信任网络(Deep Belief Nets(DBNs))<sup>[20]</sup>、Deep Boltzmann Machines(DBMs)等深度神经网络, 就是通过一种称为“Restricted Boltzmann Machines, RBMs”的简单神经网络逐层叠加构造出来的。当然在叠加的基本形式上也引进了很多变化形式, 形成了众多的计算模型。RBMs 网络的作用是在数据中寻找稀疏表达, 多层叠加则看成是对前面得到的稀疏表达的再抽象过程。所以深度神经网络完成的是从原始数据的高维空间向抽象低维认知空间的映射。最初深度神经网络属于无监督方法, 但后期的发展在网络中引入了反馈通路, 并且在参数修改过程还可以引入监督。

深度神经网络这种特征选择性和抽象性逐步增强的过程是和视觉通路对视场信息的逐级处理类似的。如在人类视觉系统中, 存在着 LGN-V1-V2-V4-IT 区视觉通路。这一视觉通路在进行视场信息处理时就完成了一个从底层特征提取到特征抽象, 再到目标识别的过程<sup>[23-25]</sup>。LGN(外膝体)的主要作用是中转来自于视网膜的神经信息, 但对信息中的特征已经呈现出轻微的选择性。V1 区接受 LGN 传来的视网膜的神经信息, 从中判断在特定位置上是否有特定朝向的边缘、特定颜色存在等。提取边缘、颜色等底层特征是这一脑区的主要任务。V2 对更加复杂的时空结构, 特别是时间结构进行反应。V4 区的特征选择性和抽象化进一步增强, 神经元会对中等复杂的模式产生反应, 如带有颜色的星形等<sup>[18]</sup>。当信息到达 IT 区后, 神经元的反应不再和目标细微特征、空间位置相关, 而是对有意义的目标整体或者组成部分做出反应。例如, 当视场中出现人脸时, 某些神经元将被激活。这些神经元的感受野几乎是整个视场, 神经元的激活将不依赖于脸的具体空间位置, 空间位置这一特征在逐步抽象化过程中被去掉了。

视觉层中这一抽象过程对于听觉、触觉等信息通路也是适用的。深度神经网络就具有这种逐步

抽象特征的能力. 以利用深度神经网络进行目标识别为例, 在最初几层的神经网络中, 提取的是简单特征, 如不同朝向的边缘等. 而随后逐层神经网络可以看成是对前一层神经网络得到的特征的再组合, 每层组合都比上一层的组合更加抽象一点, 神经元的“感受野”也更加大一些, 能够识别更大的区域内更加复杂的空间结构(目前深度神经网络用于时空间结构识别的研究还很有限). 层次越高, 能够识别的空间模式的复杂度和空间大小都会提升, 而这些模式和需要识别的目标也将更加贴近. 如识别一只猫的过程, 当输入的图像中有一只猫时, 经过表达的逐层抽象会最终激活最后一层神经网络的某路输出, 从而指示视场中的目标是一只猫. 并且随着层数的增加, 其抽象出来的特征对于目标的区分能力也是逐渐增强的, 那些和要识别的目标相关性不强的特征在抽象过程中会被逐渐弱化甚至去除. 最后一层神经网络的输出就是识别结果, 这个结果是融合倒数第二层抽象出来的特征得到的(很可能是非线性的). 当中间层的某些在训练过程中稳定出现的特征在处理新的数据时没有出现, 只要其他特征足够多或者足够强, 那么这些缺失的特征对最后的结果影响不大或者没有影响. 例如, 当视场中猫的颜色、部分轮廓外形、条纹等局部特征不一致, 甚至视场仅有一部分猫时, 依然可以鲁棒的完成对猫的识别. 这和我们的视觉经验是相符的. 漫画的理解就是一个很好的例子. 寥寥数笔就足够我们识别出所画的是哪位人物. 可见在某些典型目标识别中, 轮廓信息或者轮廓中的某些特征对于目标识别就足够了. 对于这样的目标, 深度神经网络的鲁棒性就要强得多. 但是对于仅存在细微特征差异的对象, 要将他们区分开来, 就需要更加全面、甚至是更加具体的特征.

### 3.2.2 部分深度神经网络的稀疏编码能力

RBM 网络的作用是稀疏编码. 第一个稀疏编码算法是在 1996 年由 Olshausen 和 Field<sup>[27]</sup> 提出的, 但直到近 10 年后的 2005 年才引起广泛的重视. 一个重要的原因就是当时的计算平台效率有限, 即使小规模稀疏编码算法也要耗费大量的时间, 更遑论利用这一算法去处理诸如自然场景识别等复杂任务了.

2005 年深度神经网络重新进入研究者的视野有两个原因: (1) 一个基于稀疏编码的目标识别算法在一个 101 类数据集分类中击败了所有其他的

算法, 达到了 42% 的识别精度(今天这一识别精度已经达到了 80%), 这使得人们对更为复杂的深度神经网络的性能十分期待; (2) 近年来高性能计算技术的进展, 特别是高性能计算设备价格降低, 使深度神经网络所要求的基本计算能力在很多实验室中被建立起来. 2006 年,《Science》上发表了一篇利用神经网络对数据进行降维的文章<sup>[28]</sup> 使得深度神经网络成为研究热点. 文章的主要结论是: (1) 具有多隐藏层的神经网络具有优异的特征学习和抽象能力, 提出的特征能够更加有效的和更为本质的刻画数据, 进而可以更加有效的实现信息利用; (2) 深度神经网络在训练上的难度, 可以通过“逐层初始化”(Layer-wise Pre-training) 来有效克服<sup>[28]</sup>.

### 3.2.3 深度神经网络在脑成像数据分析中的应用

深度神经网络最初应用于自然图像识别上, 因为其性能出众, 引起了各领域研究者的普遍关注. 在视觉目标识别、语音识别、自然语言理解、视频处理等多领域内得到了应用, 并显著提升了处理效率, 在个别领域内识别准确率提升了 30% 以上<sup>[21]</sup>. 2005 年起, 深度神经网络方法开始获得信息科学家的重视, 并逐渐在脑成像分析领域中获得应用并取得显著成功. 脑图像时空模式分析与自然场景识别问题有类似之处, 已经有多个研究将深度神经网络应用于磁共振功能图像序列、结构图像以及脑电信号处理中<sup>[29-30]</sup>. 在脑瘤识别中获得 83% 的识别率, 高于传统方法 10% 左右<sup>[31]</sup>. 在精神分裂症的静息态 fMRI 数据分类上, 正确识别率为 85.8%, 而目前通用的支持向量机方法得到的识别率为 76.7%<sup>[30]</sup>, 提升效果非常明显.

目前在脑成像分析领域内得以应用的机器学习方法, 其目的一般是完成数据的分类. 分类是在特征基础上完成的. 也就是说首先要使用机器学习方法从大维度的数据中提取出特征来, 然后在特征空间上完成分类. 这样的好处是分类器将面对相对于原始数据更加简单的局面, 但其缺点在于分类效果严重依赖于特征的选择, 如果特征选取不当, 分类效果可能很差. 虽然基于特征空间的机器学习分类方法取得了显著的进展, 但是近年来发展放缓, 识别率的提升都在个位数上徘徊. 深度学习直接使用原始数据, 通过学习得到稀疏化的特征, 并逐步将特征抽象化. 该方法克服了人为选择特征的主观性, 分类率提升显著, 优于目前的基于特征的机器学习方法<sup>[29]</sup>.

此外因为深度神经网络能够提取出一般机器学习方法很难得到的数据内的高维特征结构,所以适合于功能磁共振这种高空间维数以及脑电信号这种高时间维数的数据<sup>[23]</sup>。

### 3.2.4 深度神经网络实现的重要基础——高性能计算

深度神经网络参数多,训练算法复杂,训练数据往往也是超大规模的(深度神经网络和大数据是相辅相成的)。此外网络的训练具有强的主观性,需要诸多技巧和经验,例如,网络结构的选取,神经元个数的设定,权重参数的初始值选定,学习率的调整等。实际训练是一个不断试错的过程。这都使得普通计算设备无法胜任计算的要求,高性能计算成为深度神经网络面临的核心问题之一。例如,目前在图像理解、目标识别等问题中广泛应用的卷积神经网络,一般由 10~20 层构成,神经元间有几百万个连接,数以亿计的权重系数需要通过训练确定。2 年前,这种规模的训练需要数周时间,而计算技术的进步使得这一时间降到了数小时<sup>[23]</sup>。但问题是,这样的连接规模相较人脑还有不小的距离。神经网络规模的迅速增加,必将给高性能计算技术的发展速度提出高的要求。

为了提升计算效率,目前主要有三种方法<sup>①[32]</sup>: (1) 使用诸如 GPU(Graphic Process Units)等硬件计算单元进行加速。GPU 能提升计算效率的原因在于矢量化编程,提升了矩阵运算的速度。矢量化编程强调单一指令并行操作多条相似数据,形成单指令流多数据流(SIMD)的编程泛型。深层模型的算法,如 BP, Auto-Encoder, CNN 等,都可以写成矢量化形式。然而,在单个 CPU 上执行时,矢量运算会被展开成循环的形式,本质上还是串行执行,所以效率要低很多。GPU 的众核体系结构包含几千个流处理器,可将矢量运算并行化执行,大幅缩短计算时间。得益于 GPU 众核体系结构,程序在 GPU 系统上的运行速度相较于单核 CPU 往往提升几十倍乃至上千倍。一块 GPU 卡可相当于数十甚至上百台 CPU 服务器的计算能力,因此 GPU 已经成为深度神经网络模型训练方面的主要解决方案;(2) 并行计算<sup>[28]</sup>。主要包括数据并行和模型并行。数据并行是指对训练数据分组,同时采用多个模型实例对分组数据并行训练。在训练过程中,模型参数相互交换以提升收敛速度。模型并行是将模型拆分成几个相对独立的单元,几个单元并行训练;(3) 搭建 CPU 集群。利用大规模分布式计算集群的强大计算能

力,利用模型可分布式存储、参数可异步通信的特点,达到快速训练深层模型的目的。目前结合 GPU 计算和集群计算技术,构建 GPU 集群正在成为加速大规模深度神经网络训练的有效解决方案。GPU 集群搭建在 CPU-GPU 系统之上,采用万兆网卡或 Infiniband 等更加快速的网络通信设施,以及树形拓扑等逻辑网络拓扑结构。在发挥出单节点较高计算能力的基础上,再充分挖掘集群中多台服务器的协同计算能力,进一步加速大规模训练任务。

对于超大规模的深度神经网络,一般实验室的计算资源是难以胜任的,目前从深度神经网络规模角度来衡量,网络公司是超大规模的深度神经网络研究和应用的主力军。2012 年,Google 公司的 Google Brain 项目,利用一个包含了 16 000 个 CPU 核的并行计算平台训练一个超过 10 亿个神经元的深度神经网络,Google 在语音识别、图像检索等领域内的进展都和这个深度神经网络有着联系<sup>[33]</sup>。2013 年,百度成立了深度学习研究所,将深度学习应用于语音识别和图像识别、检索以及广告 CTR 预估(Click-Through-Rate Prediction, pCTR)。

脑科学研究产生的大数据集为深度神经网络的构建提供了条件。不过目前针对脑成像数据构建的深度神经网络其复杂度还是有限的,计算也主要是依靠单个实验室的计算能力完成的。如果要搭建超大规模的深度神经网络,一方面需要联合多中心的数据以支持网络的训练,另一方面也要借助高性能计算设备完成实际计算。

### 3.2.5 皮层学习算法

皮层学习算法(Cortical Learning Algorithm, CLA)也是一种人工神经网络<sup>[34]</sup>,是由 Hawkins 和他的团队提出的<sup>[35]</sup>。Hawkins 是有广泛影响力的“On Intelligence”专著<sup>[18]</sup>的作者,这本书对人工智能的发展进行了有价值的思考。CLA 算法体现着人脑皮层神经网络的“序列预测”原则<sup>[17,34-35]</sup>。在深度神经网络中,一个神经元的计算主要体现为输入信号的线性加权,或者是对输入信号一种综合。而 CLA 中的神经元有着更为复杂的动力学。通过模拟突触,神经元间可以形成抑制作用,神经元的输入可以来自于前向通路的神经元,也可以来自于侧向连接的神经元<sup>[21,34]</sup>。更为重要的是除了学习空间模式外,神经元还可以学习输入信号中的“时间模

① 深度学习及并行化实现概述. <http://djit.qq.com/article/view/1245>

式”,利用这些时间模式,神经元可以对未来输出做出预测.有了“序列模式”的预测能力,只要网络的规模足够大,神经网络就可以从经验里面学习和记忆大量的“因果链”,从而使得网络具备越来越强的处理实际问题的能力<sup>[17]</sup>.

Hawkins 建立了名为“Numenta”的公司,专门从事皮层学习算法的推广应用活动.公司于 2005 年建立,最初公司是研究性质的,2011 年,在取得突破后,公司业务开始转向应用.算法目前在金融预测、医疗领域实现应用<sup>[17]</sup>.目前此算法还没有在脑影像学中得到应用,但是因为 fMRI 等本身具有时序信息,再加上算法本身是对脑皮层神经网络更加细致的模拟,可以预见不久,皮层学习算法也会在脑研究领域内得到应用.

### 3.2.6 脉冲神经网络

脉冲神经网络(Spike Neuron Networks, SNN)被称为第三代神经网络<sup>[36]</sup>.相比于皮层神经网络,第一、二代神经网络是写意画,SNN 则是写实派的作品.这种人工神经网络的建模思路和参数选择均来自于神经科学家获得的关于脑皮层神经网络的知识,如对神经元间的连接—“突触”的信息传递过程,神经元脉冲发放的时间过程、目标的特征捆绑问题等等.相对于其他人工神经网络,SNN 在通用性、并行计算、硬件实现、计算速度、计算量等方面具有优势<sup>[36]</sup>.

神经科学家认为神经信号的信息包含在脉冲的发放频率和发放时间上,而不在脉冲的具体波形上.皮层神经网络的信息处理和传递过程表现为脉冲序列在特定神经通路上按特定时间顺序传递(在频域内表现为在特定时间段内按照特定频率传递脉冲).这种传递有两个最为重要的因素:(1)不同神经元发放频率的变化.一个神经元被激活,表现为相对于其未被激活时,脉冲发放频率增加或者减少.如对于兴奋性的神经元,如果接收到了适当的输入刺激,就会增加单位时间内输出端的脉冲发放,这时这个神经元就被激活了.相反的,对于抑制性神经元,如果在输入端得到适当的刺激,其输出段的脉冲发放频率就会下降;(2)不同神经元发放频率变化在时间上的关系,具体可以细分为“先后关系”和“同步关系”.神经元被激活仅仅意味着神经元开始参与信息处理过程,如果要保证整个过程有序有效进行,还要求不同神经元发放频率的变化在时间上满足先后关系或者同步关系.例如在人脑皮层视觉信息处理通道上,前一环节视觉区神经元发放频率的变化在

时间上会领先于其后环节神经元发放频率的变化.这种关系就是“先后关系”.又如,视觉系统的目标识别过程中,如果视觉皮层在视场中同时提取到多种符合已有先验知识的目标的特征时,那么就认为视场中就存在这个目标.也就是说目标的不同特征同时激活了负责处理不同特征的神经元群,这些神经元群在输出端同时发生了脉冲发放频率的变化,那么我们的认知系统就会认为提取到的这些特征是来自于同样一个目标,而这些特征的组合就意味着某样东西出现在了我们的视场中.相反,如果负责处理不同特征的神经元群发放频率的变化在时间上不同步,说明这些特征不是来自于同一个目标,我们的视觉系统就不会将这些特征综合起来认知<sup>[37]</sup>.

SNN 对神经元发放频率的变化以及这些变化的时序关系进行了细致的物理建模和数学建模,以期使 SNN 获得皮层神经网络信息处理的能力.对于第一、二代神经网络,一个神经元的计算过程表现为对当前时刻所有输入端数值进行加权求和.这些“权值”就是神经网络训练的核心关注点.当前时刻的“和值”和发放阈值共同决定着神经元的输出.而对于 SNN,从不同输入端得到的不再是数值,而是随机达到的脉冲串.神经元的计算过程也不再表现为实时的对这些脉冲串的信号幅值进行运算,而是在特定的时间窗口(Spike Time, 发放时间)内对这些脉冲串中包含的脉冲个数进行计算,或者说是脉冲发放频率进行计算.神经元根据发放时间内不同输入通道传递来的神经发放频率加权值来确定其输出的脉冲串的频率变化趋势.显然根据脉冲的个数而不是脉冲的幅值或者波形进行计算,运算结果要鲁棒得多.这和数字计算与模拟计算的关系有异曲同工之处.此外,神经元综合一段时间内的输入信息而不是当前时刻点上的输入信息进行运算,进一步放宽了对神经网络精确性的要求,脉冲串不必精确无误的进行传递,而容许一定的相位偏差.毕竟,大脑皮层神经网络本身就不是一个处处讲求精确的信息处理系统.在 SNN 中“先后关系”和“同步关系”是通过不同神经元各异的发放窗口间的相互配合来实现的.如两个神经元的发放窗口存在先后关系,那么即使这两个神经元得到的输入脉冲串是相同的,他们的输入依然是“先后关系”.类似的,只有当多个神经元的发放窗口是同步的,那么他们的输出才可能是同步的.通过对发放窗口的合理安排,就可以让神经元群按照希望的秩序进行工作.皮层神经网络神经元发放的“先后关系”和“同步

关系”也是通过发放窗口的选择来实现的<sup>[38]</sup>。具体过程如下:皮层中的局部场电位信号包含着大量的振荡成分,这些振荡频率最高达到 200 Hz 附近,最低达到 0.03 Hz 附近。局部场电位可以调节神经元的发放阈值,一个神经元所在区域的局部场电位信号增强,那么神经元的发放阈值就会提高,当阈值提升到相当高的水平时,发放就变得十分困难,可以认为,此时神经元进入静息状态。反过来,当局部场电位信号变低时,神经元发放阈值也会变低,相对频率较低的脉冲串输入也可能引发神经元的发放。这时,可以认为神经元进入了活跃期。局部场电位变低的这个时间段就是神经元的发放窗口。皮层中的两个区域如果它们的主导局部场电位信号频率相同,相位相同时,两个区域内的神经元激活存在“同步关系”,如果相位不同,则两个区域内的神经元激活存在“先后关系”。

从信号处理的角度来看,神经元完成的是输入和输出的映射关系,决定一个神经元的激活和发放频率有多种因素,除了输入端的脉冲频率和发放窗口这两个因素外,基因、蛋白表达、离子通道、递质类型等因素对于神经元发放也存在影响<sup>[39]</sup>。已经有工作人员对这些因素进行了研究建模,这些模型被称为“发放模型, Spike Neuron Model”。目前已经有“阈值-发放模型”、“积分-发放模型”、“考虑泄漏的积分-发放模型”、“Izhikevich 模型”等等。不同模型考虑的因素是不同的。在具体网络建立过程中需要权衡选择。这些模型将决定着 SNN 的非线性特性。在应用层面上, SNN 主要的价值有两个方面。一是皮层神经网络的模拟工具,二是在人工智能领域内的应用。Maass<sup>[22]</sup>在理论上证明了 SNN 可以一致逼近任何连续函数,这使得 SNN 在模式识别等领域内的应用前景明朗化。这个工作也成为 SNN 发展的重要里程碑。SNN 目前已经在交通路标识别、EEG 数据处理等多个领域内得到应用。目前大量研究者依然对 SNN 在理论和应用方面进行深入探索,对皮层神经网络的模拟也会随着新知识的不断积累而不断取得进展。

### 3.3 类脑计算微处理器和计算架构

目前深度神经网络等类脑计算面临的主要问题之一是模型构建、参数训练等需要的计算能力不足。这使得算法的设计和训练都变的非常耗时。虽然大规模的 CPU 集群和 GPU 等硬件系统可以显著提升目前架构的计算速度,但一方面提升后的速度依然不足以保证应用领域研究的发展速度,另一方

面这样的计算资源毕竟是有限的,仅有大型研究计划或者大公司才能提供。这显然会形成技术垄断以及随之而来的发展放缓的结果。

2012 年 IBM 公司利用当时最快的超级计算机“红杉”模拟人脑神经元网络<sup>[40]</sup>,模拟出了 5300 亿个神经元和 137 万亿个神经突触,达到了人脑神经元网络的规模,但信息处理能力依然仅有人脑的 1%<sup>①</sup>。如果利用“红杉”这样的计算机去实现人脑计算能力,需要 12 千兆瓦的功率,是纽约和洛杉矶两座城市的耗电量的总和<sup>[16]</sup>。所以利用传统微处理器和计算架构实现人脑的计算能力是不可行的。必须实现像人脑那样的高效低耗的计算方式。显然实现这种计算方式,我们也必须从微处理器和计算架构两个方面展开努力。

如果能够在芯片级别上集成类脑计算的算法和结构体系,毫无疑问将推动类脑计算走向更为广泛的应用。专用芯片在计算能力、小型化、低功耗等多方面都将超越目前的计算方式,直接推动大数据时代的机器学习。类脑芯片可以通过模拟人类大脑,以极低的能耗实现信息的异步、并行、可变速和分布式处理,彻底改变目前海量数据处理效率低下的局面<sup>[16,41]</sup>。

因为神经元信息处理方面表现出来的丰富动力学特性,使得在硬件层面上模拟神经元变得非常的困难。所以在最初的研究上,在构建神经网络的时候走的基本是软件模拟的技术路线。软件模拟在计算效率和功耗上都存在明显的不足,特别是当神经元的规模不断增加时,这些不足越来越限制了类脑计算的发展。实际上,研究者从未停止硬件模拟的研究。目前在硬件层面上对神经元的模拟取得了显著的进展,如忆阻器芯片的诞生,就为硬件模拟神经元非线性动力学和学习能力奠定了基础。忆阻器的电阻阻值可以根据流经电流的大小和方向而改变,这一阻值可改变的电阻被用来模拟神经元突触的可塑性,而突触可塑性是皮层神经网络具备学习能力的基础。利用忆阻器大量集成的机器学习芯片可以在硬件层面上模拟皮层神经网络,以此构造类脑计算微处理器。这类芯片在利用历史数据进行学习的过程中,对于相对于当前时刻较近的数据可以予以加权,突出它们在训练中的作用,这使得整

① Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH. <http://www.braininitiative.nih.gov/2025/index.htm>, 2014

个芯片具备更强的更有弹性的学习能力,更像一个皮层神经网络。

忆阻器的概念在 20 世纪 70 年代就出现了,到 2008 年才研制成功,当月就受到 DARPA 的名为“神经形态自适应可塑可扩展电子系统”项目的支持,最终目的是研制出百万神经元级的类脑微处理器。项目包含“类脑体系结构”和“神经突触核心”两个重点<sup>[16]</sup>,也就是从“计算架构”和“微处理器”两个方面开始研究。IBM 公司是“计算架构”部分的主要负责人。2014 年公司推出了共有 100 万神经元和 2.56 亿突触的“真北”类脑微处理器,有 4096 个内核,能耗不到 70 毫瓦。“真北”类脑微处理器的理论基础是 SNN,它是 SNN 的硬件实现。“真北”的计算速度是笔记本电脑的 100 倍而功耗仅有其千分之一。这样的芯片能够通过连接进行扩展,模拟更为庞大的神经网络。休斯顿实验室是“神经突触核心”部分的负责人。他们的神经突触实体采用了模拟设计,利用“时分复用”技术可以使一个硬件突触具备模拟上万突触的信息处理和传递的能力。2014 年,该实验室研制出包含 576 个神经元和 7.3 万突触的芯片。在接下来的研究阶段,IBM 和休斯顿实验室将联合将新型微处理器和新型计算架构上取得的成果集成,构建具有实时学习能力的物理计算平台<sup>[16]</sup>。

美国还设立了和类脑计算有紧密联系的若干个研究计划。美国 BRAIN 计划将对脑信息处理机制的基础科学问题展开研究<sup>①[17,42]</sup>。此计划将为信息处理专家提供更多的可以借鉴的脑信息处理方式方法方面的知识。又如“半导体技术先期研究网络计划”,其目的是提前 12 年攻克影响下一代半导体技术长期发展与应用的重大关键技术。发展新型半导体器件使其能效相对目前提升 10 000 倍以上;发展类脑认知片内微架构,将计算机性能提升若干倍。斯坦福大学的“神经栅格”项目实现了 100 万神经元和数十亿突触的实时模拟,速度是传统计算机的 9000 倍,能耗仅有其 4 万分之一<sup>[16]</sup>。

有研究者在微处理器层面上实现了深度学习神经网络等大规模人工神经网络的构建与硬件加速。如中国科学院计算所“计算机体系结构国家重点实验室”研制的深度学习系列处理器<sup>[43]</sup>。目前包括 DianNao、DaDianNao、PuDianNao 和 ShiDianNao 四种型号,分别对应四种不同的体系结构。DianNao<sup>[44]</sup>处理器的平均性能超过主流 CPU 核的 100 倍,但功耗仅为 CPU 核的 1/30~1/5,平均性能与 NVIDIA

K20M (是当前深度神经网络主要 GPU 硬件加速平台之一)相当,但功耗仅为后者的百分之一量级。DaDianNao<sup>[45]</sup>是 DianNao 的升级版,将处理器规模从 1 个提升到 16 个,同时增加了片上存储。单芯片性能超过 NVIDIA K20M 的 20 倍左右,能耗仅为其 1/300 左右。DaDianNao 主要面向大规模人工神经网络的构建和硬件加速。ShiDianNao<sup>[46]</sup>处理器则是专门针对卷积神经网络设计的,有效提升了基于卷积网络的深度学习神经网络的计算效率。能耗为主流 GPU 的 1/60 左右,但计算效率是其 30 倍以上<sup>[46]</sup>。

虽然目前各国正在实施的类脑计算相关项目取得了显著的进步。项目也描绘出了似乎是触手可及的美好未来,但是我们距离“像大脑一样计算”还相当遥远。无论是传统的微处理器及其架构,还是目前像“真北”、“DianNao”那样的新型微处理器和计算架构,其计算能力和计算功耗和人脑相比还有明显差距。我们构造的可实现的神经网络的规模距离人脑规模还有 1000 倍以上的差别。之所以有这样巨大的差距,一方面是我们现有的模拟脑计算的能力非常有限,另一方面是我们距离通晓人脑智能所有方面的目标还有相当大的差距。如果依据现有知识搭建系统,我们还需要一个又一个的假设,一个又一个的简化,以至于模拟本身可能实现的只是神经科学家对大脑的理性推测而不是大脑本身。以欧洲的“人脑工程”为例,在计划开始后就争议不断,甚至近 600 位科学家联名给欧盟写信反对目前计划设定的目标和技术路线<sup>②[47]</sup>。他们中的相当一部分认为目前实现全脑级的模拟从知识储备和技术储备来看时机都还不成熟。对于美国脑研究计划也存在类似的批评,认为计划提出的在短时期内实现对大规模(上百万)神经元的实时无损观测是不现实的<sup>③[42]</sup>。但相比较而言,美国的 BRAIN 计划比欧盟的 HBP 计划设定的目标现实,具有相当的可行性<sup>④[47-48]</sup>,因而受到的批评要小些,没有遇到像 HBP 那样公开大规模抵制的情况。

- ① Dharmendra S M. Introducing a brain-inspired computer: TrueNorth's neurons to revolutionize system architecture. <http://www.research.ibm.com/articles/brain-chip.shtml>
- ② Simple I. Scientists threaten to boycott 1.2 billion human brain project. the Guardian, monday 7 July 2014. <http://www.theguardian.com/science/2014/jul/07/human-brain-project-researchers-threaten-boycott>
- ③ Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH. <http://www.braininitiative.nih.gov/2025/index.htm>, 2014

## 4 讨 论

### 4.1 脑科学研究进入规模化、协作研究阶段, 研究的复杂性要求高性能计算能力持续提升

人类大脑有 860 亿个神经元, 这些神经元彼此通过短程和长程神经纤维直接或者间接的联系在一起, 构成一个迄今为止所知的最为复杂的神经网络. 这些神经元间连接的全貌对科学界依然是个谜, 我们只是在若干具体脑区、若干层次上对他们有了一个一般性的理解.

从 20 世纪 90 年代开始, 脑科学研究进入了快速发展阶段, 发达经济体都迅速增加了对脑科学研究的资助力度. 脑观测手段也取得了长足的进步, 出现了磁共振、正电子断层扫描、近红外成像、微电极阵列、膜片钳等众多目前成为主力手段的观测技术. 电生理、脑电波等前期的观测手段, 其性能也出现了质的变化. 计算技术的快速发展则解决了不断积累的脑观测数据的分析需求. 种种有利条件, 使得众多研究者开始进入脑科学的研究领域. 在国内, 1985 年在钱学森的推动下我国建立了思维科学学科, 1993 年成立的香山科学会议确定了赶超世界脑科学研究的总体目标. 到 2010 年, 国内知名的研究机构几乎都成立了脑科学研究机构(中心). 不同研究团队根据各自的兴趣与研究条件选择了不同的关注点, 在不同层次、不同侧面展开了对脑的解读. 大体上研究可以分成三个层次: 微观、介观和宏观. 在微观层面, 主要研究的问题有: 神经元各离子通道递质的传递与接受、突触信息传递的非线性特性、神经元对兴奋性和抑制性输入的综合过程、神经元长短程连接的原则和具体细节等等. 在介观层面, 主要研究的问题有: 相似功能神经元聚集的原则和其信息传递与处理的具体细节、不同神经元群的连接原则和其蕴含的信息处理机制、兴奋性和抑制性神经元群的相互作用、感觉通道上不同脑区的合作机制以及他们在信息处理过程中的作用等等. 在宏观层面上, 主要研究的问题有: 脑功能网络、记忆、情感等问题. 因为对象复杂性, 同一个研究团队一般只能选择一个层次上的研究, 并且非常可能还需要进一步的限定研究范围. 此外, 脑科学研究平台昂贵, 即使一流团队也只能拥有若干个研究手段, 从有限视角对选定的问题进行探究. 如此一来, 虽然近 30 年脑科学研究空前繁荣, 取得了丰富的研究成果, 很多领域内

的进展甚至是突破性的, 但是整体看来, 即使汇总目前得到的所有知识, 依然无法回答脑科学的终极问题: “智能是什么? 从何而来?”. 显然智能不是从某个特定层面、特定视角获得的, 而是多层次、多侧面协同作用获得的. 如果要理解大脑, 不但要积累局部、细节的知识, 也要在一个统一的框架里面全面的组合这些知识. 早在 20 世纪 80 年代, DNA 结构发现者, 诺贝尔奖得主佛朗西斯·克里克就曾指出神经科学的研究虽然使人类获得了大量关于脑的知识, 但是这些知识只是一堆没有任何理论的庞大数据<sup>[49]</sup>. 今天, 这种情况还没有得到根本的改变. 我们需要寻找一个统一的框架去组合、解释我们手头的知识, 这样才有可能在整体上理解人类智能, 也才有可能在理解的基础上去更好的模拟或者利用人脑智能. 目前来看要组合这些知识, 要构建统一的框架, 不是个别研究团队或者小规模的研究联盟可以解决的, 我们需要的是大规模的、协作式的研究机制.

脑科学研究者已经意识到进行系统研究的重要性. 目前美国、欧盟以及我国即将颁布的脑科学研究计划都将理解智能作为首要目标, 并且依据这一目标制订了系统的研究计划. 众多研究团队, 特别是顶尖的研究团队将花去相当多的研究时间, 根据一个统一设定的路线图开展研究. 可以期望的是, 我们将得到更加一致的、全景的知识网络. 在这一新的知识积累的过程中, 研究者面临的将是海量数据, 这是因为: (1) 研究对象的生理机制和动力学过程异常复杂, 使得观测数据集庞大; (2) 单种观测手段性能提升带来的数据量的增加; (3) 针对同样的对象, 多种观测手段同时积累数据; (4) 确保结果可靠需要采集大样本数据集. 这使得研究对计算平台的要求和依赖性都显著提升. 个人计算平台、中小型计算平台不能支撑其计算的需求, 目前的高性能计算平台也只能支撑部分研究. 欧盟和美国的脑计划中也都有关于用于脑科学研究的高性能计算平台的建设计划. 如欧盟的脑计划, 启动至今不过 2 年时间, 已经有 4 台超级计算机用于研究工作. 即使如此, 因为计算的限制, 目前的很多研究还只能采用简化的方法进行, 以降低对计算的要求. 例如, 2015 年欧盟 HBP 计划中的研究团队发表在 *Cell*<sup>[8]</sup> 上的文章来看, 使用 4 台世界排名前 100 名的超级计算机, 也只能完成对幼年大鼠运动感觉皮层的局部区域的数值模拟. 并且这种模拟依据的模型是经过简化的. 2012 年 IBM 公司利用当时最快的超级计算机“红杉”模

拟人脑神经网络,虽然神经元数量级达到了人脑水平,但是因为模型过于简化,模拟出来的神经网络仅有人脑的计算能力的1%<sup>①</sup>。如果以目前我们对人脑神经网络的了解来完全建模人脑,其数字化模拟是不可行的。所以高性能计算还必须经历一个迅速的成长期,并且必需持续进步,才可能完全支撑未来各脑研究计划的顺利展开。

#### 4.2 大脑高性能计算能力的本质是智能

我们说“计算机的计算能力远低于人脑的计算能力”,指的不是单位时间内代数运算的次数,而是在解决实际问题时,人脑表现出来的高度智能。对于那些我们可以轻易说出结果,但是无法对其过程进行详尽描述的认知过程,对于计算机来说都是非常困难,甚至是目前无法完成的任务。例如一个高明的指挥员面对复杂的战争态势,可以准确判断出敌方意图,并迅速给出对策。经验丰富的医生通过简单的观察即可确定一些需要经过复杂检测确定的病症。

计算机在代数运算速度、信号传递速度、误差控制、存储能力等方面是优于人脑的。在数字计算机系统中信号的传递速度是光速,而在大脑中,信号在神经纤维里的传输速度是“米/秒”级的。电脑的每个存储计算单元犯错的比率非常之低以至于在实际应用中可以不予考虑,神经元在传递信号时并不可靠,相同刺激下每次传出的脉冲频率和波形都有变化。在进行计算时,计算机的误差很小并且可控,在绝大部分的应用中可以忽略,而神经元群在提取信号特征时存在显著的误差,如在人脑视觉初级视觉皮层 V1 区内,神经元“计算”出的光栅朝向误差可能高达 30°。人脑能够记忆的数据量很小,并且还存在着遗忘过程,计算机的存储能力却几乎是无限的。显然在单纯的“计算”环节,人脑是完败的。但问题是,当计算机用于具体问题的计算时,使用的方法是固定的,是人脑“理性思维”设计的解决方法,而不是人类也说不清楚的人脑自身的解决方法。这些方法即使巧妙,也往往摆脱不了“蛮力计算”的形象。这使得大量的计算资源都消耗在这些低效的计算方法上。人工智能试图解决的问题之一就是计算方法的低效问题。它希望通过算法设计,使得计算机具备人类智能的某些特征,从而使计算机具备人脑智能。应该说目前人工智能已经取得显著进展。人工智能有两个基本思路:一是考察人类思维过程,从中抽取一些信息处理原理和规则,在信息处理过程中使用这些原理规则提升处理效率,如专家系统等。二是建立类似人脑

神经网络的大规模神经网络。最近的深度神经网络是截至目前这种思路最高的发展形态。在第一种思路指导下建立的方法取得成功是领域性的和对象依赖的,或者说必须针对特定的领域、特定的目的,设计特定的方法。第二种思路指导下建立的方法在初期并不成功,早期的人工神经网络一般仅有三层,最多数百个神经元。也只是用于有显著约束问题的解决,如手写字母识别。直到深度神经网络出现后这类方法才表现出显著的优势,和人脑神经网络类似,深度神经网络的基本结构是稳定的,经过训练既可用于解决不同问题,如语音识别、交通路标识别。不过深度网络的训练本身是需要高性能计算的支撑的,也有研究开始在硬件上模拟神经元群的功能,形成了所谓的类脑芯片。这类芯片的模拟级别在百万级,并且神经元的功能是经过简化的。利用这些芯片搭建的计算平台在特定应用背景下表现出了相当的智能化水平和低功耗特性。

依据冯·诺依曼计算架构的传统高性能计算提升空间已经有限。下一步的发展方向有量子计算,也有类脑计算。其中的类脑计算就是在软硬件两个方面模拟人类大脑神经网络。特别是当神经网络的规模相当于或者远超过人脑级别后,有望在数量级上提升计算效率并同时扩展应用领域,最终会使得这类计算架构的计算平台成为一种通用平台。并且这类平台在解决实际问题时需要的代数运算量会出现数量级的下降。所以在类脑计算这个思路下,高性能计算关注的是智能,是计算效率,而不再单纯是代数计算速度的提升。

目前在软硬件上对大脑的模拟,依据的都是某种简化后的模型。原因有:(1)我们还没有得到完整的神经元、神经元群、微回路、功能区以及功能网络的全景图,目前的知识链存在缺环。在记忆、情感等问题上积累的知识还相当有限;(2)因为计算能力或者硬件规模的限制,模拟需要必要的简化。如目前即使使用世界一流的超级计算机,也无法将我们得到的关于人脑神经网络的知识在人脑级别的神经网络上予以实现(或者这种实现在时间上不可行);(3)必要的简化是合理的。人脑在构造的底层,特别是神经元级别上并不是一个像钟表一样精

① Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director, NIH. <http://www.braininitiative.nih.gov/2025/index.htm>, 2014

密运转的单元,存在很多不精确性和不完善的地方。至于其生化和电学的动力学行为中,哪些是智能必须的,哪些是神经元功能和结构不完善的表现,我们还没有搞清楚。如人脑的某些生理学、形态学特征可能是大脑进化过程中的阶段性结果,这些就不应该成为模拟的对象。所以完全的模拟也是不合理的。神经元行为本身的不精确性也说明智能的产生可能并不需要一丝不苟的底层单元,简化是可取的。问题是简化到什么程度?这个问题需要在未来类脑计算和脑科学本身的发展中逐步予以回答。

#### 4.3 模拟人脑智能还需要实现抽象思维能力

无论是利用软件模拟人脑的深度神经网络还是利用硅基电路模拟人脑的“真北”类脑芯片,模拟的都是人脑灰质皮层的视觉通路特别是视觉通路的神经网络。不严格的说,这些类脑计算方法虽然一定程度上模拟了人脑的视觉等神经通路的信息处理机制,但是他们还不具备模拟人脑“思维”的能力。人脑具备“智能”,能够“思维”,除了拥有规模巨大的感觉通路神经网络外,还因为人脑拥有“抽象思维/概念思维”这一关键能力。例如,在视场中出现了一只猫,这一输入经过逐级处理,最终会激活 IT 区的某些神经元。这些神经元的激活说明视觉系统从背景中“区别”出了猫,但这时猫并没有进入“意识”。IT 区神经元的激活还必须进一步激活“海马”区负责概念存储的神经元,让“猫”这一概念出现在“意识”中。这时猫的认识才告完成。如果希望进一步完成“这只猫是不是经常出现在花园中的那只猫”这种更为复杂的识别任务,还需要神经通路中“前馈”和“反馈”通路进一步的复杂配合。当“猫”这一概念进入意识后,人脑就可以利用这一“概念”联想出很多东西,如旧日和猫有关的经历,还可以思考如“猫是否会抓伤自己或者传播某些疾病”这样的问题,进而思考出相应的规避行为。人脑可以摆脱实体进行概念层面的学习记忆和逻辑思维,并且能够以现有知识作为素材通过逻辑思维获得新知。这样的学习是“永无止境”的。就是因为人脑具有“抽象思维”的能力,才使得人脑具备了相较于其他动物强得多的智能。目前类脑计算技术还不具备“抽象思维”的能力。无论是人工神经网络还是类脑计算芯片,完成的都是从输入到输出的映射,也就是说信息处理必须是外部输入发起的。他们无法像人脑一样进行“内源性”思维。

人工智能中有一大类方法是利用规则进行符号

推理,这样的方法具备人脑这种抽象思维的某些特征。但是这类符号推理因为其规则是人为定制的,推理的联想和发散能力较弱,所以不能像人脑一样有效应对新情况和学习新的知识。概率机器学习方法(Probabilistic Machine Learning),如贝叶斯推理等,从数据中获得的经验,并将经验用于新数据的处理,这种方法使用经验去外推的能力有限,也就是如果遇到的情况和之前的情况有显著差别时,对于经验的使用就没有人脑那样灵活。这类学习方法可以很好地建模不确定性,但对于不确定性的建模并不是在所有的问题中都是重要的。例如,对于目标识别、语音识别、上下文理解等领域,如果存在大量样本,深度神经网络的效果要好于贝叶斯推理等方法。深度神经网络并没有明确考虑不确定性因素,说明对不确定性建模在大样本下并不是至关重要的。但是对于样本稀少,获得推理的前提存在诸多不确定性时,概率学习方法还是有其优势的<sup>[50]</sup>。在类脑计算的研究中,也应该融合基于概率进行学习的思路。贝叶斯推理的原则就被认为是人类以及动物通过经验进行学习的方式<sup>[51-52]</sup>,目前已经有研究讨论皮层神经环路是如何实现贝叶斯推理的<sup>[53-54]</sup>,可能的研究成果可以用于目前人工神经网络的改造,以使其具有更强的学习能力。

实现高效的类脑计算,不见得仅依靠一种技术,可以考虑多种技术的结合,如将深度神经网络和基于贝叶斯推理的人工智能方法结合<sup>[50]</sup>。深度神经网络学习数据中的各种非线性特征,输出高维向量,基于这些高维向量,利用贝叶斯推理最终得到结果。目前的贝叶斯推理主要是基于符号的,这样就需要对贝叶斯推理进行改造使其适用于基于高维向量的推理。

## 5 结束语

大数据时代的脑科学研究面临着前所未有的发展机遇。借助新技术的研发与相关研究领域取得的关键性突破,如神经网络绘图工具、高时空分辨率 MRI 扫描仪的研制以及纳米科学等领域的蓬勃发展,人们对于自身大脑的认识越来越深入。

高性能计算技术在脑科学研究中起着越来越重要的作用,不论是进行大数据的存储与分析的云计算技术,还是“人脑计划”中脑模拟平台的搭建,都要运用高性能计算技术提供的大规模数据处理与高精

度仿真计算能力;同时,在基于信息通讯技术的新型脑研究模式下,高性能计算与脑科学研究紧密结合、相互促进,由此派生出的高性能计算的智能应用,将会对未来的科学、经济与社会产生极其深远的影响。

类脑计算技术的发展将为紧凑型、低能耗系统的研发开辟道路. 超级计算机虽然能给我们带来强大的计算能力,但核心的高速运转也带来了巨大的功耗,而人类大脑的功耗却处于极低水平. 因此,以人脑的工作原理指导未来计算技术的发展,让计算机模拟人脑的运行机制,超级计算机以及其他大功耗设备的能耗问题将有望在未来得到解决。

脑科学与高性能计算技术的结合将成为探索信息处理高效能的有效途径. 超级计算机的运算速度能够达到千万亿次,但是其智能水平极为低下,若能将计算机的高性能与人的高智能相结合,那么在大数据时代借助这样的“机器脑”,则有望大幅提升数据挖掘的效率;同时,相关的理论研究成果也会对深度神经网络等这类学习算法的研究产生帮助,从而更好地指导如机器视觉、语音识别、大规模机器翻译等技术的发展<sup>①</sup>。

## 参 考 文 献

- [1] Zhu Li-Jun, Zhu Yuan-Gui, et al. The funding and prospect of global brain research. *Science Foundation in China*, 2013, 27(6): 359-362(in Chinese)  
(朱丽君, 朱元贵等. 全球脑研究计划与展望. *中国科学基金*, 2013, 27(6): 359-362)
- [2] Vivien M. The big challenges of big data. *Nature*, 2013, 498(7453): 255-260
- [3] Amunts K, Lepage C, et al. BigBrain: An ultrahigh-resolution 3D human brain model. *Science*, 2013, 340(6139): 1472-1475
- [4] Hawrylycz M J, Lein E S, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 2012, 489(7416): 391-399
- [5] Jiang T, Liu Y, et al. Multimodal magnetic resonance imaging for brain disorders: Advances and perspectives. *Brain Imaging and Behavior*, 2008, 2(4): 249-257
- [6] Van Horn J D, Toga A W. Multisite neuroimaging trials. *Curr Opin Neurol*, 2009, 22(4): 370-378
- [7] Breakthrough of the year, Areas to watch. *Science*, 2012, 338(6114): 1528-1529
- [8] Markram H, Muller E, Ramaswamy S, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell*, 2015, 163(2): 456-492
- [9] Big data. *Nature*, Special online collection, 2008
- [10] Dealing with data. *Science*, Special online collection, 2011
- [11] Liu Run-Sheng. The U. S. Government proclaimed the launch of the human brain project. *Scientific Chinese*, 2013, (9): 18-19(in Chinese)  
(刘润生. 美国政府宣布启动人脑计划. *科学中国人*, 2013, (9): 18-19)
- [12] Kou Da-Zhi. A number of concerns in high performance computing fields—ISC'11 sidelights. *Development and Applications of High Performance Computing*, 2013, 35(2): 12-13(in Chinese)  
(寇大志. 高性能计算领域的若干关注点——ISC'11 侧记. *高性能计算发展与应用*, 2013, 35(2): 12-13)
- [13] Luo Shu-Qian. Research of the digital brain atlas based on human brain specimen. *Chinese Journal of Medical Instrumentation*, 2001, 25(2): 91-94(in Chinese)  
(罗述谦. 数字化人脑图谱研究. *中国医疗器械杂志*, 2001, 25(2): 91-94)
- [14] Liu Run-Sheng, Zhang Li-Juan. Ambitious EU human brain project. *Scientific Chinese*, 2013, (9): 16-18(in Chinese)  
(刘润生, 张丽娟. 雄心勃勃的欧盟人脑项目. *科学中国人*, 2013, (9): 16-18)
- [15] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks. *Proceedings of the Advances in Neural Information Processing Systems*, 2014, 4(27): 3104-3112
- [16] Wang Wei. Europe and America actively developing new brain like microprocessors. *China Integrated Circuit*, 2014, 23(11): 87-90(in Chinese)  
(王巍. 欧美积极发展新兴类脑微处理器. *中国集成电路*, 2014, 23(11): 87-90)
- [17] Jonathan L. From neural networks to deep learning: Zeroing in on the human brain. *XRDS*, fall, 2011, 18(1): 29-34
- [18] Hawkins J. *On Intelligence*. New York, USA: Times Books Press, 2004
- [19] Roe A W, Pallas S L, et al. Visual projections routed to the auditory pathway in ferrets: Receptive fields of visual neurons in primary auditory cortex. *The Journal of Neuroscience*, 1992, 12(9): 3651-3664
- [20] Hinton G E, Osindero S, et al. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, 18(7): 1527-1554
- [21] Hawkins J, George D, et al. Sequence memory for prediction, inference and behavior. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 2009, 364(1521): 1203-1209
- [22] Maass W. *Computing with Spiking Neurons in Pulsed Neural Networks*. Cambridge, MA: The MIT Press, 2001: 55-85
- [23] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, 521(7553): 436-444

① Hernandez D. The man behind the Google brain: Andrew Ng and the quest for the new AI. *Wired Magazine*. <http://www.wired.com/2013/05/neuro-artificial-intelligence>, 2014

- [24] Felleman D J, Essen D C V. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1991, 1(1): 1-47
- [25] Cadieu C F, et al. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLOS Computational Biology*, 2014, 10(12): e1003963-e1003963
- [26] Zhang T L, Zeng Y, Xu B. HCNN: A neural network model for combining local and global features towards human-like classification. *International Journal of Pattern Recognition and Artificial Intelligence*, 2016, 30(1): 1655004
- [27] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, 381(6583): 607-609
- [28] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 28(313): 504-507
- [29] Plis S M, Hjelm D R, et al. Deep learning for neuroimaging: A validation study. *Frontiers in Neuroscience*, 2014, 8(8): 229-229
- [30] Kim J, Calhoun V D, et al. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage*, 2016, 124(A): 127-146
- [31] Gladis Pushpa Rathi V P, Palani S. Brain tumor detection and classification using deep learning classifier on MRI images. *Research Journal of Applied Sciences, Engineering and Technology*, 2015, 10(2): 177-187
- [32] Dean J, Corrado G S, et al. Large scale distributed deep networks//Proceedings of the Neural Information Processing Systems (NIPS'12). Lake Tahoe, USA, 2013: 1223-1232
- [33] Le M A, Ranzato R, et al. Building high-level features using large scale unsupervised learning//Proceedings of the International Conference on Machine Learning, Edinburgh, Scotland, 2012: 8595-8598
- [34] Ahmad S, Hawkins J. Properties of sparse distributed representations and their application to hierarchical temporal memory. Subutai Ahmad, 2015, arXiv: 1503.07469
- [35] Hawkins J, George D. Hierarchical temporal memory. Alphascript Publishing, 2010, Suppl(5): 1-10
- [36] Falah Y H, Ahmed B Y, Haza N A H. Computing with spiking neuron networks—A review. *Applied Soft Computing*, 2014, 6(1): 1-21
- [37] Gray C M, Konig P, Engel A K, Singer W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 1989, 338(6213): 334-337
- [38] György B. Rhythms of the Brain. Kettering Northants, USA; Oxford University Press, 2006: 373-432
- [39] Kasabov N. Integrative connectionist learning systems inspired by nature: Current models, future trends and challenges. *Natural Computing*, 2009, 8(2): 199-218
- [40] Gu Fan-Zhi. Recent developments of the EU and US brain researches. *Science*, 2014, 66(5): 16-21(in Chinese)  
(顾凡之. 欧盟和美国两大脑研究计划之近况. *科学*, 2014, 66(5): 16-21)
- [41] Zeng Yi, Liu Cheng-Lin, Tan Tie-Niu. Retrospect and outlook of brain-inspired intelligence research. *Chinese Journal of Computers*, 2016, 39(1): 212-222(in Chinese)  
(曾毅, 刘成林, 谭铁牛. 类脑智能研究的回顾与展望. *计算机学报*, 2016, 39(1): 212-222)
- [42] Theodore M, Wong R P, et al. IBM Research Report, RJ10502(ALM211-004). San Jose, USA, 2012
- [43] Chen Yun-Ji. The hardware of neural network in view of a researcher in computer architecture. *China Computer Federation Bulletin*, 2015, 11(7): 10-22(in Chinese)  
(陈云霁. 体系结构研究者眼中的神经网络硬件. *中国计算机学会通讯*, 2015, 11(7): 10-22)
- [44] Chen T S, Du Z D, et al. DianNao: A small footprint high-throughput accelerator for ubiquitous machine-learning//Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems. Salt Lake City, USA, 2014: 269-284
- [45] Chen Y J, Luo T, et al. DaDianNao: A machine-learning supercomputer//Proceedings of the 47th Annual IEEE/ACM International Symposium on Microarchitecture(MICRO-47). Cambridge, UK, 2014, 17(2): 112-121
- [46] Du Z D, Chen Y J, et al. ShiDianNao: Shifting vision processing closer to the sensor//Proceedings of the 42nd ACM/IEEE International Symposium on Computer Architecture. Oregon, Portland, 2015: 92-104
- [47] Yuste R, Church G M. The new century of the brain. *Scientific American*, 2014, 310(3): 38-45
- [48] Underwood E. Brain project draws presidential interest, but mixed reactions. *Science*, 2013, 339(6123): 1022-1023
- [49] Crick F H C. Thinking about the brain. *Scientific American*, 1979, 241(3): 181-188
- [50] Zoubin G. Probabilistic machine learning and artificial intelligence. *Nature*, 2015, 521(7553): 452-459
- [51] Knill D, Richards W. Perception as Bayesian Inference. Cambridge Town, UK; Cambridge University Press, 1996
- [52] Griffiths T L, Tenenbaum J B. Optimal predictions in everyday cognition. *Psychological Science*, 2006, 17(9): 767-773
- [53] Doya K, Ishii S, et al. Bayesian Brain: Probabilistic Approaches to Neural Coding. Massachusetts, USA; MIT Press, 2007
- [54] Deneve S. Bayesian spiking neurons I: Inference. *Neural Computation*, 2008, 20(1): 91-117



**LIU Ya-Dong**, born in 1977, Ph.D., associate professor. His research interests include neuroimaging, brain-computer interface, image/signal processing.

**HU De-Wen**, born in 1963, Ph.D., professor. His research interests include cognitive science, brain-computer interface, and neural networks.

## Background

This paper discusses the relationship between high performance computing and brain science researches. High performance computing and brain science are two research fields that are current rapidly developing and continuing to make breakthroughs. In recent years, the areas of the two studies begin to blend; they provide each other the research inspiration and methods. High performance computing has become a vital means of brain research, which opens up new research perspectives and patterns for it, such as ‘Big data theology’, ‘Cloud computing theology’, ‘Supercomputing theology’, et al.. In turn, brain research gives rise to the emergence of a modern computing pattern by putting forward new requirements for high performance computing, and has inspired several promising theology which will improve effect, efficiency, power consumption, et al. of high

performance computing to a brand new height. These promising theologies include Deep learning, Cortical learning algorithm, Brain-inspired CPU Chip, et al. This trend will continue, their link will be even more closely in the future. This article aims to discuss the current intersection of these two research fields and their possible fusion configuration in future. The authors’ research interests include neuroimaging, brain-computer interface, neural networks, et al. With their excellent work in the field of brain science, the authors achieve the Second Prize of National Natural Science Award and the First Prize for Natural Science Award of Ministry of Education. This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61473305 and 61375034, and in part by the National Basic Research Program of China under Grant No. 2013CB329400.