

《异步脉冲神经网络》

+

《神经拟态计算系统》

人工智能汇集

国内外人工智能跟帖留言

(共124条跟帖)

点评：弱人工智能向强人工智能转型 异步脉冲神经网络+神经拟态计算系统

陆首群

2022.4.11

研发基于异步脉冲神经网络的神经拟态计算系统，实现由弱人工智能向强人工智能转型之路，在这方面表现突出的有：

英特尔研制的 Loihi 芯片，由 Loihi 芯片集成的 Pohoiki Springs 神经拟态网络（由 1 亿神经元组成），及神经拟态计算系统，2019 年已制成原型机，由英特尔团队合作开发应用生态。

由曼彻斯特大学团队合作研发由 10 亿个神经元组成的神经拟态网络，2018 年 11 月已制成 Spin Naker 类脑计算机原型机。

浙江大学自主研发达尔文-2 芯片，由芯片集成 1.2 亿个神经元组成的神经拟态网络，2020 年 9 月发布已研制成功 Darwin Mouse 类脑计算机原型机，研发工作正在继续跟进。丹麦科技大学与亚琛科技大学合作也在研发基于脉冲神经网络的神经拟态计算系统（包括探索建立神经拟态学习规则的库，多模态神经脚本知识模型等）。

下面详细介绍它们的研发工作：

1) 英特尔公司

2017 年研发 Loihi 第一代芯片，制程 14nm，单颗芯片 13 万个神经元

2021 年研发 Loihi 第二代芯片，制程 7nm，单颗芯片 100 万个神经元。

在推出第二代神经拟态研究芯片 Loihi-2 的同时，推出用于开发神经启发应用的开源软件框架 Lava。

Lava 软件框架是一个开放、模块化、可扩展的框架，能够在跨越异构架构上无缝运行，实现跨平台执行，以及各种人工智能、神经拟态和机器人框架的互操作性。

第一代 Loihi 芯片使神经拟态处理速度较之传统计算机提高 1000 倍，能耗下降 10000 倍；第二代 Loihi 芯片和 Lava 框架极大地提高了神经拟态处理速度，较第一代 Loihi 芯片又提高了 10 倍（为传统计算机运算速度的 10000 倍）。

2019 年 3 月 19 日，英特尔将 768 颗 Loihi-1 芯片集成，组成 Pohoiki Springs 神经拟态网络（由 1 亿个神经元组网），拥有平行处理性能和异步信号传输能力。

2019 年制成基于 pohokisprings 神经拟态网络的 1 亿神经元的类脑计算机原型机。

脉冲神经网络（SNNs）是神经拟态计算中一种全新的模型，可以模仿人脑中自然神经元网络方式，将计算模块重新分布，神经拟态在芯片上计算、存储单元一体化，颠覆传统冯诺依曼的架构，实现类脑的智能功能，神经拟态计算被认为是引领下一代人工智能的主要计算模式。

神经拟态计算未来主要的应用前景，一是基于神经拟态计算开展超算应用开发，二是用于非结构化数据、实时的场景，三是用于多模态的、实时的场景（如机器人、无人机需要持续学习、自适应学习的场景）。为了扩大应用，英特尔于 2020 年 3 月成立神经拟态研究社区（INRC），

推动神经拟态算法、软件和应用程序的研发,参加 INRC 的除英特尔外,还有 IBM 、HP、MIT、普渡大学、斯坦福大学、以及联想、埃森哲、罗技、梅赛德斯奔驰、机器视觉传感器公司 Prophesee,还有罗格斯大学、戴尔施特理工大学等。

2) 曼彻斯特大学

自 2016 年始,曼彻斯特大学与南安普顿大学、剑桥大学、谢菲尔德大学、ARM、Silistix 公司、泰雷兹等合作共同研发脉冲神经网络(芯片),并在芯片集成的基础上研发神经拟态计算系统(和类脑计算机)。

每颗第一代 SpinNaker-1 芯片支持 1.6 万个神经元,曼大将 SpinNaker-1 芯片集成: $30k \times 16k = 4.6$ 亿个神经元,由第二代 SpinNaker-2 芯片集成为 10 亿个神经元。由 10 亿个神经元组成的神经拟态网络,并在此基础上研发 SpinNaker 类脑计算机(神经拟态+并行处理),原型机于 2018 年 11 月制成对外发布。

3) 浙江大学

2015 年自主研发达尔文-1 芯片,2019 年自主研发达尔文-2 芯片(每颗芯片 576 个内核,每颗芯片支持 15 万个神经元),2019 年 8 月,浙江大学联合之江实验室发布由 792 颗芯片集成的神经拟态网络(由 1.2 亿个神经元、近千亿个突触组网)。

2020 年 9 月发布我国首台基于亿级神经元组成神经拟态网络,并由此研制而成的 Darwin Mouse 类脑计算机原型机,功耗为 350-500w,由三个高为 1.6m 的标准机柜组成。

该原型机已在：(1) 抗洪抢险场景下支持、调控机器人协调工作，(2) 建立脑外侧膝状核的神经网络模型，(3) 建立学习-记忆融合模型。

该机运行机制与 SpinNaker 相似，研发应用场景正在起步。

从上面 3 例看来，首先设计拟态芯片，实行芯片集成，生成 1 亿个-10 亿个神经元，组成亿级神经拟态网络，支持研制神经拟态计算系统（即类脑计算机），制成原型机，颠覆传统的冯诺依曼架构，形成神经拟态计算（类脑计算）架构，随即组成开发应用生态系统（英特尔公司、曼彻斯特大学两例均集结一批合作伙伴共同开发）。

国内外人工智能跟帖留言共 124 条

1-2, 脉冲神经网络 (SNN) 可以模仿人脑的电气信号, 是模拟生物神经元连接和运行方式的模型, 通过计算产生神经电脉冲进行信息传递。目前, 国内外一些专家正在对脉冲神经网络进行研究, 未来人工智能的发展期待对脉冲神经网络和类脑算法的关键性突破。

2-8, 2017 年清华大学施路平教授等研发“世界首款双控异构融合类电脑芯片” (将基于脉冲神经网络 SNN 的类脑计算算法与基于人工神经网络 ANN 的深度学习算法集成到一颗芯片“天机芯 (Tianjic)”上, 将人工智能的应用创新推进到更接近于具有“自主思考”成份的“认知智能”阶段。第二代“天机芯 (Tianjic)”的照片刊登在英国《nature》杂志的封面上。

3-9, IBM 从 2014 年开始研发 True North 芯片, 进行底层理论研究 (即对类脑脉冲神经网络 SNN 及类脑算法的研究), 以期推动人工智能发展。两年前 IBM 发布类脑芯片 True North 引起了轰动。IBM 先搞了一个简易的核心电路, 采用 asynchronous circuit, 当时是非常大胆的, 这个研究项目主要是为了验证神经形态计算硬件。因为芯片面积超过 100 平方毫米, 良品率非常低, 不能量产, 随后 Synapse Project 已结束了。

4-10, 脉冲神经网络类脑芯片的进步将会带来计算体系结构的革命, 是下一代人工智能发展的重要突破方向。由浙江大学牵头研发的脉冲神经网络类脑芯片“达尔文 2”已于近期发布 (2015 年发布“达尔文 1”, 现在已启动研究“达尔文 3”),

达尔文 2 单芯片由 576 个内核组成，每个内核支持 256 个神经元、神经突触 1000 多万（即单芯片支持的神经元规模达 15 万个），经过系统扩展可构建千万级神经元类脑计算系统（在神经元数目上相当于果蝇的神经元数目），达尔文 2 是目前已知单芯片神经元规模居全国前列。该类脑芯片目前主要应用于图像物体识别、视频音频理解、自然语言处理、脑电识别，较之基于深度学习的人工神经网络，具有独到优势，更擅长于模糊数据处理。

5-13，纯脉冲神经网络超越深度网络才是更大看点，我们（指国内）正在做，希望一两年能有突破。

6-14，脉冲神经网络不是 IBM 首创，TrueNorth 当时的卖点是芯片实现了大规模脉冲神经网络、低功耗。

7-15，清华这辆自行车成精了，可自我平衡，自动绕障避障，自主管控，自主决策。

8-25，脑神经科学成果是人工智能源头活水。

9-26，稀疏矩阵乘法、张量运算（张量图）恐怕是主要类脑设备实现。

10-29，清华大学研发了世界首款双控异构融合类电脑芯片，将基于生物脉冲神经网络的类脑算法与基于人工神经网络的深度学习算法结合起来；另外，也需要探索将知识推理和数据驱动结合起来，以推动人工智能的发展。

11-33，转发刊于《智东西》的6款类脑芯片：① IBM TrueNorth 芯片，② Intel Loihi 芯片，③高通 Zeroth 芯片，④西井科技 Deep South 芯片，⑤浙大“达尔文”类脑芯片，⑥AI-CTX 芯片。

12-34-6，清华大学在通用人工智慧能上开发了新型人工智能芯片——天机芯（TianJic），这是一款基于脉冲神经网络（SNN）的类脑算法，与基于人工神经网络（ANN）的深度学习算法集成于一块芯片上。

13-43，生物学灵感，神经网络代表着一种受人类大脑启发的信息处理范例。在大脑中，神经元将轴突和树突紧密连接，并通过突触在它们之间传递化学信号。人类大脑拥有约 1000 亿个神经元，每个神经元最多与 10,000 个其他神经元相连。

14-46，采用生物神经网络（SNN）的类脑算法是否比采用人工神经网络（ANN）的深度学习算法提高了一个档次？！

这里提出的类脑算法的机制和形式是什么？在某些情况下，类脑算法是否也可以深度学习算法来表达？请教专家。

15-55，英特尔研究院院长 Richard (Rich) A. Uhlig 谈在数据洪流中属于颠覆性技术的新型计算方式：量子计算、神经拟态、图计算、概率计算，其中神经拟态芯片模仿人脑运作机制，形成高能效神经网络系统，主要采用异步脉冲神经网络（SNN）去解决问题，可应用在机器人、网络、动态控制、稀疏编码、图式搜索、路径规划、约束满足等很多领域。

16-63, 每个神经元激励 fire 的阈值大约 0.07 伏, 属于电气工程数量级 (一节 AA 电池 1.5 伏), 计算机与人脑增强交互应该很快推动人工智能进步。

17-150-7, 冲突爆发: 符号主义和连接主义的古老争论再燃:

以 Twitter 为阵地, 以加里·马库斯(Gary Marcus)为首的长达一年的争论, 为围绕人工智能发展方向数十年的争论注入了新的活力。

马库斯是纽约大学教授、作家、企业家, 以及基于逻辑的 AI 的鼓吹者, 他发起了不懈的 Twitter 争论, 试图打破深度学习的根基, 并推广其他人工智能方法。

Gary Marcus 马库斯重新点燃了所谓的符号主义者和连接主义者之间的古老争论, 前者坚持认为基于规则的算法对认知至关重要, 而后者则认为将足够多的神经元与正确的损失函数连接起来是获得机器智能的最佳途径。

与马库斯针锋相对的 AI 从业者重新熟悉了象征主义的方法, 以免连接主义的局限性导致资金崩溃, 或人工智能陷入寒冬。这一争论促使人们对人工智能的未来做出了清醒的评估, 并在 12 月 23 日由马库斯和深度学习先驱、蒙特利尔大学 Yoshua Bengio 教授的一场现场辩论中达到高潮。辩论过程非常有礼貌, 双方都承认两个党派之间需要合作。

2018 年 12 月, 马库斯向深度学习支持者的“帝国主义”态度发起了挑战, 开始了自己的进攻。他继而鞭策 Facebook 的深度学习先驱 Yann LeCun, 要他选择一方: 是把自己的信仰寄托在纯粹的深度学习上, 还是有好的“出色的老式人工智能”(good old-fashioned AI)的一席之地?

OpenAI 在 10 月份提出的混合模型成为头条新闻。它的机械手通过深度强化学习和经典的 Kociemba 算法的结合解决了魔方难题。虽然马库斯指出是 Kociemba 算

法计算出了解决方案，而不是深度学习，但其他人断言机器人可以通过进一步的训练来学习这项技能。

去年 12 月，微软提出“神经符号人工智能” (neurosymbolic AI)，填补了这一空缺。这是一个旨在弥合神经表示和符号表示之间差距的模型架构。随着 2019 年临近结束，NeurIPS 会议强调了人工智能社区的 soul searching。谷歌研究员 Blaise Aguerre y Arcas 在一次主题演讲中表示：“我们目前所有的训练模式都是为了让 AI 在特定任务中取得胜利或者赢得高分，但这并不是智能的全部。”

符号主义者和连接主义者之间的敌意可以追溯到半个多世纪以前。1969 年，马文·明斯基和西摩尔·派普特在《Perceptrons》书中，仔细分析了以感知机为代表的单层神经网络系统的功能及局限，证明感知机不能解决简单的异或 (XOR) 等线性不可分问题，帮助触发了第一个 AI 冬天。

第二个 AI 寒冬是在将近 20 年后，部分原因是符号 AI 依赖于 LISP 计算机，而 LISP 计算机已经随着 PC 的出现变得过时了。

神经网络在 20 世纪 90 年代开始普及，并在过去十年计算能力和数据的爆炸式增长中取得了主导地位。当连接主义者和符号主义者齐头并进，或者直到一个派别消灭另一个派别时，我们期待着激动人心的新时代。

18-151，新一代人工智能研究方向：（1）研究新一代人工智能基础理论（机理、模型和算法）；（2）研发面向需求的共性技术（以神经网络和算法为核心、数据和硬件为基础、数据知识融合驱动为动力、应用落地为目标，形成开放兼容稳定成熟的技术体系）；（3）开发高端、低耗、廉价的智能芯片和系统；（4）建立人工智能开源开放的基础平台。

19-165, 提高了神经网络的准确性和效率(采访波士顿大学计算机科学系统 AI 专家)

神经网络体系结构将继续扩大规模和深度, 并产生更准确的结果, 并将更好地模仿人类在涉及数据分析的任务上的表现。同时, 提高神经网络效率的方法也将得到改善。

随着 AI 进入越来越多的领域, 将受到更多的可靠性和偏见的审查, 例如越来越多的地方政府出于隐私和公平考虑, 禁止使用 AI 技术进行监视。

20-210, 按大数据建立起来的深度学习 AI 系统是不可理解、不可解释的。溯根求源: 大脑是模拟的, 机器(计算机)是数字的, 深度学习将数字输入大脑, 大脑是不识别的, 而大脑中模拟的思维逻辑传到机器, 机器也是不识别的, 所以采用深度学习算法技术难以做到脑机交互, 而深度人工智能如果得不到大脑支持, 将在暗箱操作中变得不可理解、不可解释, 难以创造如同人类的自主思维、创意和情感。

认知计算是用来构建模拟人类思维过程的系统。

知识图谱是实现认知智能的关键, 有望成为“大脑”。它基于知识表示进行逻辑推理, 真正把数据驱动和知识驱动及知识推理结合起来, 知识图谱是一种大规模语义网络(或自然语言处理 NLP)。

2019 年 NLP 取得了飞跃式发展, 提高了语言转变(如从模拟讯号转变为数字讯号)的精准度(使大脑思维的模拟讯号变成机器能够识别/理解的数字讯号), 从而推动计算机建模。

同时又在专门语料库上利用知识增强语义理解框解推动对 NLP 的微调, 进一步提

高语言转变的精准度。

但机器最大的问题是缺乏常识，而建立常识库又非常困难，这个问题仍然阻碍了机器/计算机对自然语言的理解。

IBM Watson 在探索医疗人工智能认知计算过程中就遇到了缺乏常识带来的困扰，在痛定思痛后，他们提出了具人 (embodiment) 概念，必须与现场临床医生沟通、取得共识、统一行动，以弥补研究人员常识之不足 (将基础理论和常识结合起来)，才能完成医疗防治的任务。

21-273, 人工智能与人脑智能

近年来人工智能取得了巨大进步，受人脑神经元调节机制启发的人工智能新颖算法层出不穷！但人工智能离人脑智能（或称人类智能）还差得很远，而人脑（神经元）自身的开发远未充分！

22-305, 基于生物脉冲神经网络 (SNN) 的类脑算法现状和趋向

从网上看到随《评人工智能如何走向新阶段》一文后，国内外的跟贴留言至今已达 300 多条，热烈得很！

人们对人工智能发展前沿的研发动态感兴趣，我们非常赞成编辑部为了解除大家顾虑，围绕“公开”的动态进行转发和讨论。

选题如当前国内外采用基于生物脉冲神经网络的类脑算法的研发前沿其动态如何？

现在我转发几位热情的朋友马上作出的介绍：

请阅读 2016 年 11 月出版的《中国图像图形学报》由黄铁军（北大）、施路平（清

华)、唐华锦(川大)、潘纲(浙大)、陈云霄(中科院计算所)、于俊清(华中科大)等6位AI专家联合撰写的文章《类脑计算的研究进展与发展趋势》;也请大家阅读在中国计算机学会技术报告(2015~2016)中还是由上述6位作者撰写《类脑计算的研究进展与发展趋势》的文章。

我在这里还转发另一位朋友介绍的文章:

发表在《Science》杂志2020年2月号(122卷)上由Aboozar Taherkhani, T.M.McGinnity两位作者撰写文章《A review of learning in biologically plausible spiking neural networks》,这篇文章很好的对SNN现状进行总结。

23-310, 神经形态计算一直被寄以厚望, 它将宣告摩尔定律的终结

研究者开发了一种模拟生物嗅觉的神经形态算法, 这是模拟或突破神秘大脑迈出的一步, 代表可以开发出当前人工智能趋势的新方法

神经形态计算可以大幅度提升数据处理能力和机器学习能力

英特尔神经形态计算实验室科学家纳比尔-伊姆艾姆与康奈尔大学生理学实验室研究员托马斯-克莱兰德率领的人工智能联合研发团队研发出一种用来模拟生物嗅觉的神经形态算法, 这项成果于3月16日发表在英国《自然-机器智能》杂志上。

他们采用英特尔开发的Loihi神经形态芯片, 描述一种基于哺乳动物嗅觉系统的神经形态算法, 通过神经脉冲或尖峰传递信息并自动调节突触强度, 他们研发过程中的难点是透彻理解并攻克对生物神经网络的层面所实现的算法。

他们利用苯、氨、丙酮、一氧化碳和甲烷对其进行气味训练。

24-311, 探索生物脉冲神经网络理想的训练方法

当前基于传统神经网络的机器学习/深度学习算法让我们在人工智能很多领域实现了突破,但这些神经网络在生物学上是不够精确的,它们不能模仿生物大脑神经元的运作机制。

脉冲神经网络(Spiking neural network, SNN)将脉冲神经元作为计算单元,能够模仿人类大脑的信息编码和处理过程;通过脉冲或尖峰传递信息,通过由神经元和突触组成的网络,使用拟合生物神经元机制的模型进行计算,通过学习和训练提供强大的计算能力,从而建立类脑算法,把人工智能推向强人工智能的新阶段。

目前国内外不少 AI 专家正在研发 SNN,但 SNN 还处于早期研发阶段,一些露头的 SNN 仅能处理相对简单的任务。因此除要对拟合生物神经元机制的模型需继续进行深入的研发外,对目前尚不成熟的 SNN 的训练方法也需进行深入的研发探索。由于神经元复杂的动态性及不可导操作,暂时还没有一个可扩展的训练方法。近期采用较多的 SNN 训练方法是 DNN—to—SNN (将目标 DNN 转化),效率不高,到实用尚有不少距离,需要加快研发!

25-313, 1 亿神经元大规模的神经拟态硬件计算系统诞生

今年 3 月 19 日英特尔宣布研制出神经拟态计算系统 PohoikiSprings,由 1 亿个神经元组成,集成到 Loihi 芯片上,相当于一个小型哺乳动物的大脑。

与本集跟贴留言介绍的、集成生物脉冲网络 SNN 的芯片比较:

IBM North 芯片(2014),由 100 万个神经元、2.56 亿个突触组成(神经回路);

清华大学天机芯(TianJic)由 4 万个神经元、1000 万个突触组成;浙江大学达

尔文—2 芯片（2018）由 15 万个神经元组成。

英特尔于 2017 年开发了代号为 Loihi 的第一款自主学习神经拟态芯片，由 1000 个神经元硬件设计架构（模拟多个逻辑神经元）组成，与训练人工智能系统同级通用计算芯片相比，Loihi 芯片（用新颖异步脉冲方式计算）能效提升 1000 倍。2019 年英特尔在 Loihi 基础上推出 Pohoiki Beach（64 块 Loihi 芯片，800 万个神经元）；2020 年 3 月 19 日又推出 Pohoiki Springs（768 块 Loihi 芯片，1 亿个神经元），该神经拟态系统拥有超级并行和异步信号传输能力，在明显降低功耗同时显著提升性能。

人类大脑由 800~1000 亿个神经元组成，每个神经元有很多突触，构成神经回路（每个神经元和其他神经元之间会有连接，通过突触构成回路），人类所有知识就储存在这些连接或回路里面。

研发大规模神经拟态（神经元）计算系统，也是研发基于生物脉冲神经网络类脑算法超前的、基础性的、重要的研究步骤。

26-317，对深度脉冲神经网络 SNN 的训练是一大挑战

对 SNN 训练可能是从神经元层面理解大脑运行方式的最可行方法

作者: Amirhossein Tavanael (Louisiana 大学)、Masoud Ghodrati (Monash 大学)、Saeed Reza Kheradpisheh (Kharazmi 大学)、Timothée Masquelier (CNRS 大学)、Anthony Maida (Louisiana 大学)。

这篇论文发表在《Neural Networks》，111 卷，2019 年 3 月

该论文描述了以监督和无监督学习方法，训练生物脉冲神经网络 (SNN) 在准确性

和算力成本上的比较，研究发现（当时的）SNN 在准确性上仍不及人工神经网络（ANN），但差距正在缩小，要指出的是 SNN 通常操作次数比 ANN 少很多，在处理时空数据方面是更好的选择。

该论文被大量引用。

该论文摘要：深度学习在近年来发展迅猛，尤其在计算机视觉场景硕果累累。这一场景通常对深度（多层）人工神经网络（ANN）进行监督学习-反向传播的训练，这一训练方式需要大量的标注数据，收获很高的分类准确程度。

ANN 中的神经元是单独的、静态的、连续激活的，而生物神经元则使用离散的脉冲进行计算与信息传递，该过程的效果取决于脉冲频率和脉冲时间。所以从生物的可实现性来说，脉冲神经网络 SNN 要高于 ANN。生物神经元的脉冲在时间和空间上的分布是松散的，由事件驱动。如果结合合适的本地学习规则，我们将能针对 SNN 建造更低功耗的仿生硬件。但是，深度 SNN 的训练是一大挑战。脉冲神经元的传递函数通常是不可微分的，从而无法反向传播。

27-318，清华大学自主研发第二枚类脑芯片再次登上《nature》杂志

基于多阵列忆阻器存算一体系统其处理能效优于 GPU

今年 2 月 27 日清华大学钱鹤、吴华强教授率领的 AI 研发团队，成功研发出一款基于多阵列忆阻器存算一体系统，搭载在一颗神经形态计算/类脑芯片上，在处理卷积神经网络（CNN）时其能效比前沿的图形处理芯片（GPU）高两个数量级。

28-319，通过人工智能算法模仿人类大脑

DeepMind 研究多巴胺系统向通用人工智能进军

DeepMind 正在学习、研究人脑中的多巴胺神经递质系统：如何激活大脑，如何理解在神经元连接中的大脑运行方式，如何生成意念、动机、欲望、创意等感觉，建立新的神经网络，开发类脑算法，迈向通用人工智能。

29-321，MIT 研发 EIG 深度神经网络模型

可供探索人脑工作原理作参考

计算机视觉是目前人工智能最成功的领域之一，但这些用来检测物体、人脸的复杂系统，至今也无法与人类视觉系统相提并论。MIT 领导的研究小组建立一种名为 EIG (efficient inverse graphics) 的特殊深度神经网络模型，以展示神经层次结构如何快速推断出场景的潜在特征，与常用的带标签数据训练出来的神经系统相比，新模型的机理更接近真实的人类视觉系统，可以像人类一样从图像中迅速生成详细的场景描述。EIG 是从一个模型中训练出来的，这个模型反映了人脑针对所见人脸生成的内部表征。这一成果（论文）发表在《Science Advance》上，为探索大脑工作原理提供参考。

30-322，英特尔康奈尔科学家共建人工智能生物嗅觉系统

英特尔 AI 科学家 Nabil Imam 研究团队在康奈尔大学研究动物嗅觉系统的科学家帮助下共同构建一种方法：让神经拟态芯片 Loihi 通过样本训练掌握 10 种危险品不同气味的神经长征，找到一种智能的、可靠的和快速的化学传感处理系统。搭载神经拟态芯片的新方案比 Loihi 此前的传统方案，展现出更加出色的识别准确率。该系统还可应用于医疗系统、机场安检区域（以识别危险物品）。

31-345，编者按：报导“清华大学自主研发第二枚类脑芯片再次登上《nature》杂志上”的跟贴已发表在跟贴留言 318 条上，最近又收到几件重复同样内容的跟贴，谨选其中一件跟贴破例发表在 345 条跟贴留言上。即：

中国 AI 科学家频频出手！清华第二枚类脑芯片再登《自然》杂志

中国科学家又一次抢占了世界人工智能研究制高点！

2 月 27 日，清华大学钱鹤、吴华强教授团队成功研发出一款基于多阵列忆阻器存算一体系统，在处理卷积神经网络（CNN）时能效比前沿的图形处理器芯片（GPU）高两个数量级。

2020 年，AI 技术将会有哪些改变？近日，科技日报的一篇权威报道给出了答案：人工智能的下一步发展路径将会是类脑计算。

类脑计算又被称为神经形态计算，与传统深度学习相比，类脑计算优势明显。深度学习是一种机器学习方法，也是目前主流的人工智能算法。

在大数据背景下，人工智能每一项场景落地应用，都离不开海量数据的支撑。以百度自动驾驶技术为例，从前期的试验到最终场景落地，整个过程中需要进行上千亿次的数据喂养，时间耗费预计长达十年。

另外，人工智能算法训练中心在执行任务时，动辄消耗电量几千瓦甚至几十千瓦，而人的大脑耗能却仅相当于 20 瓦左右。基于此，专家们一致认为：眼下正遍地开花的深度学习可能并非人工智能的终极方案。

在产业界，美国芯片巨头英特尔最早嗅到类脑计算的商机。2019 年 7 月，英特尔发布消息称，其神经形态研究芯片 Loihi 执行专用任务的速度可比普通 CPU 快 1000 倍，效率高 10000 倍。11 月中旬，英特尔宣布：埃森哲、空中客车、通用电气和日立公司加入英特尔神经形态研究共同体（INRC），该共同体目前已拥有超

过 75 个成员机构。

但这并不意味着美国在类脑计算领域已领先一步。最早一批看到类脑计算未来的中国科学家，已成功研发了全球第一款类脑芯片！

去年 8 月，英国《自然》杂志首次刊登中国自主研发的天机芯片，清华大学精密仪器系施路平教授团队研发的类脑计算芯片“天机芯”首次公开亮相。

天机芯片可以同时支持机器学习算法和类脑电路，它由 156 个 FCores 组成，包含约 40000 个神经元和 1000 万个突触，采用 28 纳米工艺制程，面积为 3.8×3.8 平方毫米。

这款芯片未来可应用到自动驾驶和智能机器人等场景，在试验阶段，施路平教授的研究团队用一辆无人自行车验证了“天机芯”的处理能力。

从视频来看，载入天机芯的无人自行车不仅可以识别语音指令、实现自平衡控制，还能对前方目标人员进行探测和跟踪，并自动过障、避障，还展现了自适应姿态控制、语音理解控制、自主决策等功能，看上去就像是机械有了颗大脑一般。

天机芯片如其名——天机不可泄露。它的诞生并不容易，在天机芯的研究初步阶段，没有任何资料文献可供施路平教授团队参考。一个细节是，有一次为了寻找灵感与突破口，他独自登山时故意让自己走丢，从而思考怎么在困境之中找到出路。

在施路平教授的带领下，研究团队奋斗了整整七年光阴。在《自然》杂志公开选用论文之前，团队几乎没有发表过任何研究成果。也恰恰是因为没有“论文绩效”的压力，他们才得以心无旁骛地进行研究。

长江后浪推前浪。凭借自主研发对的天机芯，中国青年科学家在世界舞台已崭露头角。4 个月前，施路平教授的博士生邓磊被选为 2019 年《麻省理工科技评论》

“35 岁以下科技创新 35 人”中国区得主。

特别值得一提的是，清华大学自主研发的第二枚类脑芯片近日又登上《自然》杂志。2 月 27 日，清华大学钱鹤、吴华强教授团队成功研发出一款基于多阵列忆阻器存算一体系统，在处理卷积神经网络（CNN）时能效比前沿的图形处理器芯片（GPU）高两个数量级。

毫无疑问，在人工智能时代，中国已悄然成为领导者！

32-352, Nature 长文综述：类脑智能与脉冲神经网络前沿

2019 年 11 月 28 日普渡大学 Kaushik Roy、Akhilesh Jaiswal 和 Priyadarshini Panda 在《Nature》杂志上（Nature575, 607-617/2019）发表 Towards Spike-based machine intelligence with neuromorphic Computing

摘要：本文概述神经形态计算在算法和硬件方面的发展，介绍学习和硬件框架的原理，以及神经形态计算的主要挑战及发展前景、算法和硬件的协同设计等方面的内容。

导语：在人工智能如火如荼的今天，基于人脑的“脉冲”（Spiking）模拟计算框架下的脉冲神经网络（SNN）、神经形态计算（neuromorphic computing）有望在实现人工智能的同时，降低计算平台的能耗。这一跨学科领域以硅电路实现生物中的神经环路（circuit），现已发展到包括基于脉冲的编码以及事件驱动表示的算法的硬件实现。

目录：

一，算法展望

二，在 SNNs 中学习基于转换的方法

三，其他有待研究的方向

四，硬件展望

五，算法-硬件协同设计

六，总结

附：参考文献

33-353，创造具有类人脑的技术一直都是人工智能创新的源泉

(nature (2019.11) 类脑智能与脉冲神经网络前沿一文摘要一)

从前科学家一直认为人脑中的信息是通过不同的通道(channels)和频率传递的，如今他们认为人脑就像一台计算机。随着神经网络的发展，今天计算机已在多个认知任务中展现出非凡能力。

人脑能够执行惊人的任务(例如同时识别多个目标、推理、控制和移动)，而能耗只有近 2W，相比之下，标准计算机仅识别 1000 种不同的特体能耗就需要 250w。尽管人脑尚未被探索穷尽，但人脑非凡能力归结于三个基本观察：广泛的连通性，结构和动能化的组织层次，以及时间依赖(time dependent)的神经元突触连接。神经元(neurons)是人脑的计算原始元素，它通过离散动作电位(discrete action potentials)或“脉冲”交换和传递信息。突触(synapses)是记忆和学习的基本存储元素。

人脑拥有数十亿个神经元网络，通过数万亿个突触相互连接。基于脉冲的时间处理机制使得稀疏而有效的信息在人脑中传递。

目前最先进的人工智能总体使用的是这种受到人脑层次结构和神经突触框架启发的神经网络。现代深度学习网络(DLNs)本质上是层级结构的人造物，像人脑

一样用多个层级表征潜在特征，经过转换形成的。像人脑层次结构那样，各种硅基计算单元以层级方式排列，以实现高效的数据交换。尽管两者在表面上有相似之处，但人脑和硅基计算机计算原理之间存在鲜明区别：①计算机中计算（处理单元）和存储（存储单元）是分离的，不同于人脑中计算（神经元）和存储（突触）是一体的，②受限于二维连接的计算机硬件，人脑中大量存在的三维连通性目前无法在硅基技术上模拟，③晶体管主要为了构建确定性布尔（数字）电路开关，和人脑基于脉冲的事件驱动型随机计算不同。

但是，使得“通用智能”（包括基于云服务器到边缘设备）无法实现的主要瓶颈是巨大能耗和吞吐量需求。

在人脑指引下，通过脉冲驱动通信从而实现了神经元-突触计算的硬件系统将可以实现节能型机器智能。神经形态计算始于 20 世纪 80 年代晶体管仿照神经元和突触的功能动作，之后迅速演化到包括事件驱动的计算本质（离散的“脉冲”人造物），最终在 21 世纪初期这种研究努力促进了大规模神经形态芯片的出现。

（待续）

34-354，建立一个突触权重可调节的脉冲神经元模型

脉冲驱动型计算（SNNs）优缺点

（nature 2019.11 类脑智能与脉冲神经网络前沿 摘要二）

今天算法设计师们正在积极探索（“学习”）脉冲驱动型计算的优缺点，去推动可扩展性、高效的“脉冲神经网络”（spiking neural networks, SNN）。

我们可将神经形态计算领域描述为一种协同工作，它在硬件和算法域两者中权重相同，以实现脉冲型人工智能。谈到“智能”（算法）方面，包括不同的学习机制

（无监督以及基于脉冲的监督，或梯度下降方案），同时突出显示要利用基于时空事件的表征。算法有用于对抗硬件漏洞的鲁棒性，可以实现能耗和精度之间的最佳平衡。

算法展望：脉冲神经网络

按照神经元功能，将神经网络分为三个代际：①第一代被称为 McCulloch-Pitt 感知机，②第二代神经元单元增加了连续非线性（使其能计算一组连续输出值），③第三代神经网络主要使用“整合放电”（integrate-and-fire）型尖峰神经元，通过脉冲交换信息。

SNN 使用信号的时间（脉冲）处理信息。脉冲本质上是二进制事件（0, 1），SNNs 中的神经元单元只有在接收或发出尖峰信号时才处于活跃状态，因此它是事件驱动型的，可以使其节省能耗。若无事件发生，SNNs 单元则保持闲置状态。SNN 中的输入值为 1 或 0，也减少了数字上的点积运算，减小了求和的计算量。

神经形态工程师的一个主要目标是：在利用基于事件（使用基于事件的传感器）及数据驱动更新的同时，建立一个具有适当突触可塑性（即突触的权重调节）做脉冲神经元模型，从而实现高效识别、推理等智能应用。

SNNs 最大优势在于其能够充分利用基于时空事件的信息。今天我们有相当成熟的神经形态传感器来记录环境实时的动态改变，这些动态感官数据可与 SNNs 的时间处理能力相结合，以实现低能耗的计算。

高效的 SNNs 框架结合基于脉冲的学习规划，可以产生有效的训练。

尖峰神经元可以实现非连续的信息传递，并发出不可微分的离散脉冲，它们不能使用基于梯度下降型的反向传播技术（这是传统神经网络训练的基础）。

SNNs 的缺点是在大多数学习任务中效果仍落后于第二代的深度学习。

SNNs 还受限于基于脉冲的数据可用性。

SNNs 训练算法的识别性能是在现有静态图像的数据集上进行评估的(如 CIFAR 或 ImageNet)。

35-355, 本文总结

nature (2019.11) 类脑智能与脉冲神经网络前沿一文摘要三

神经形态计算 (一种高效方式)

脉冲神经范式的算法含义

算法-硬件协同设计的跨层优化

基于脉冲的节能智能系统

如今,“智能化”已经成为我们周围所有学科的主题,本文阐述了神经形态计算作为一种高效方式,通过硬件(计算)和算法(智能)的协同演化的方式来实现机器智能。

本文首先讨论了脉冲神经范式的算法含义,这种范式使用事件驱动计算,而不是传统深度学习范式中的数值计算。描写了实现标准分类任务的学习规则(例如基于脉冲的梯度下降、无监督 STDP 和从深度学习到脉冲模型的转换方法)的优点和局限性。

未来的算法研究应该利用基于脉冲信号的信息处理的稀疏和时间动态特性;以及可以产生实时识别的互补神经形态学数据集;硬件开发应侧重于事件驱动的计算、内存和计算单元的协调,以及模拟神经突触的动态特征。特别引人关注的是新兴的非易失性技术,这项技术支持了原位混合信号的模拟计算。本文也讨论了包含算法-硬件协同设计的跨层优化的前景。例如,利用算法适应性(局部学习)和硬

件可行性（实现随机脉冲）。

最后谈到，基于传统和新兴设备构建基于脉冲的节能智能系统与当前无处不在的人工智能相比，两者的前景其实是相吻合的。现在是我们该交换理念的时候了，通过设备、通路、架构和算法等多学科的努力，通过合作打造一台真正节能且智能的机器。

36-356，与神经科学建立联系/神经形态计算的出现

nature (2019.11) 类脑智能与脉冲神经网络前沿一文摘要四

与神经科学建立联系

与神经科学的研究成果相结合，把这些抽象的结果应用到学习规则中，以提高学习效率。如 Masquelier 等人利用 STDP 和时间编码模拟视觉神经皮层，他们发现不同的神经元能学习到不同的特征，这一点类似卷积层学到不同的特征。研究者把树突学习和结构可塑性结合起来，把树突的连接数作为一个超参数，以此为学习提供更多可能。SNN 领域的一项互补研究是 LSM (liquid state machines)。LSM 利用的是未经训练、随机链接的递归网络框架，该网络对序列识别任务表现卓越。但是在复杂的大规模任务上的表现能力仍有待提高。

神经形态计算的出现

在 20 世纪 80 年代，在生物神经系统领域，Carver Mead 设想了“更智能”、“更高效”的硅基计算机结构，他的工作代表了计算硬件领域的一种新的范式。Mead 并不在意 AND、OR 等布尔运算。相反他利用金属氧化物硅 (MOS) 晶体管在亚阈值区的电气物理特性（电压—电流指数相关）来模拟指数神经元的动力学特征。这样的设备—通路协同设计是神经形态计算中最有趣的领域之一。

37-357, “超级大脑”芯片/忆阻点积

nature (2019.11) 类脑智能与脉冲神经网络前沿一文摘要五

“超级大脑”芯片

“超级大脑”芯片的特点是整合了百万计的神经元和突触，神经元和突触提供了脉冲计算的能力。Neurogrid 和 TrueNorth 分别是基于混合信号模拟电路和数字电路的两种模型芯片。Neurogrid 使用数字电路，因为模拟电路容易积累错误，且芯片制造过程中错误影响也较大。设计神经网络旨在帮助科学家模拟大脑活动，通过复杂的神经元运作机制—比如离子通道的开启和关闭，以及突触特有的生物行为。相比而言，TrueNorth 作为一款神经芯片，目的是用于重要商业任务，例如使用 SNN 分类识别任务；而且 TrueNorth 是基于简化的神经元突触原型来设计的。

以 TrueNorth 为例，主要特征如下：

异步地址事件表示 (Asynchronous address event representation)：首先异步地址事件表示不同于传统的芯片设计（所有的计算都按照全局时钟进行），但是因为 SNN 是稀疏的，仅当脉冲产生时才要进行计算，所以异步事件驱动的计算模式更适合进行脉冲计算。

芯片网络 (networks-on-chip, NOCs) 可以用于脉冲通信，NOCs 就是芯片上的路由器网络，通过时分复用技术用总线收发数据包。大规模芯片必须使用 NOC，是因为在硅片加工过程中，连接主要是二维的，在第三维度灵活程度有限。因此尽管使用了 NOC，芯片的联通程度仍然不能和大脑中的三维连通相比。包括 TrueNorth 在内的大规模数字神经芯片，如 Loihi，已经展示除了 SNN 技术以外的应用效果。使得我们能更加接近生物仿真技术。不过，有限的连通性，NOC 总

线带宽的限制，和全数字方法仍然需要进一步研究。

忆阻点积

作为模拟计算的一个实例，忆阻点积 (Memristive dot products) 是实现原位神经形态计算的一种有前景的方法。可是表示点积的忆阻阵列中产生的电流既有空间依赖性又有数据依赖性，这使得交叉开关电路分析成为一个非常复杂的问题。研究交叉开关电路非理想状态的影响，探索减轻点积不准确影响的训练方法的研究并不多，而且这些工作大部分集中在深度神经网络而不是 SNN 中。然而我们可以合理假设，在这些工作中开发出的基本器件和电路的见解也能用于 SNN 的实现。现有工作需要精致的设备一通路模拟运行，必须与训练算法紧密耦合，以减少精度损失。基于最新设备的交叉开关阵列的理论模型，以及为点积误差建立理论边界的努力，都将引起人们的关注。这将使算法设计者无需耗时、设计迭代设备一通路一算法模拟，就能探索新的训练算法，同时也能解决硬件不一致的问题。

38-361, 基于清华“天机”芯片，“灵汐科技”可实现自行车无人驾驶

“灵汐科技”是一家类脑芯片研发商 (2018.1 成立)

该自行车能够追踪移动目标，跟随目标直行、转弯、避障，车身搭载平衡传感器确保车身不倒，根据情况灵敏调整，能够进行语音识别，通过语音指令执行转向、加速等多种操作。

“天机芯”是清华大学施路平教授团队研发的全球首款异构融合类脑芯片 (2019.8 登上 nature 封面)。

39-385, 一位院士、脑科学专家的提问

- ①未来, 可不可以充分模拟人类大脑, 研究出更高级的人工智能技术?
- ②人脑与机器能否高度融合?
- ③如果未来脑科学推动人工智能研究取得更大突破, 机器会不会和人一样有思想、有情感、有意识, 它们还能和人类和谐相处吗?
- ④目前机器与人脑相比还有多大差距?
- ⑤脑机融合(脑机接口)技术前景如何?

(2019年夏天马斯克在猴脑内植入芯片, 猴子通过猴脑成功操控电脑, 马斯克宣布取得这项技术的突破)

- ⑥脑机融合技术是否可能会成为未来人工智能的一个热门方向?

40-394, 基于 Loihi 计算架构的神经拟态计算。

2020年4月9日英特尔中国研究院院长宋继强接受记者采访, 谈到英特尔对未来计算的研究和布局问题时, 他说量子计算和神经拟态计算是非常重要的新兴计算方式。他在谈到神经拟态计算时说, 面向神经拟态计算, 英特尔发布了 Loihi 新型计算架构, 可以模拟人脑神经元连接构建的连接方式, 将计算和存储融合, 并考虑到时间序列, 采用“异步脉冲”方式进行计算。

关于英特尔 Loihi 神经拟态芯片(整合计算和存储)其主要参数: 128个内核、13万个神经元、1.3亿个突触(每个神经拟态计算内核模拟1000个逻辑神经元, 片上网络连接支持高效的脉冲消息分发, 高度复杂的神经网络拓扑, 支持多种学习模式的可扩展的片上学习能力)。

在神经拟态芯片提供的底层硬件基础上有两类做法:

- ①第一类模拟设计不变, 把现有深度神经网络移植到神经拟态芯片上, 需要大量

数据，只是运行网络的功耗更低。

②第二类涉及模型设计和算法基本原理的改变。模型设计首先需要用非深度学习的方法去做，现在配合神经拟态计算比较常用的就是脉冲神经网络（SNN）。

SNN 充分考虑了时间序列上的差异，在设计该网络时会模仿一些生物感知和处理节奏。比如英特尔近期发布的基于神经拟态芯片 Loihi 的嗅觉系统，其网络构造系统借鉴了人类嗅觉系统的结构设计。这是全新的模型设计，它不像深度学习那样需要大量数据、大量参数来达到一个稳定状态。

41-405，从神经元个体膜电位生长可变的神经动力系统。

在神经拟态工程学中，通过神经元集成建模、突触相连建网（组建大规模脉冲神经网络）、（从单个神经元膜电位）到生长可变的神经动力系统，形成脉冲（尖峰）神经网络工作机制。

美国华盛顿大学（UW）电子系统工程学院、生物工程学院的研究团队于 2020 年 4 月 7 日在 biorxiv 平台上发表了一篇文章：脉冲（尖峰）神经元和神经元集成模型及生长可变的神经动力系统。

该文介绍神经元的动态、脉冲响应、神经元集成建模，以及通过突触相连形成的大规模的脉冲神经网络，单个神经元产生的动作电位（膜电位）和其状态以及与神经元集成相应的生长可变的神经动力系统，该文介绍了一种新型的神经元集成及其脉冲神经网络模型。

该模型的优点是它可独立控制三个神经动力学特性：①控制稳态的神经元集成模型的动态，该神经元集成动态编码了精确的网络能量功能的最小值；②控制网络中单个神经元产生的动作电位的状态（不影响网络能量的最小值）；③在不影响

网络能量最小值或动作电位状态的情况下，控制脉冲统计和瞬时神经元集成动态。所提模型的核心是生长可变的神经元集成及其脉冲神经网络的动力学系统的各种变体（各种可变的神经动力系统解决方案）。

无论网络规模和神经元连接性的类型（抑制性或兴奋性）如何，均可产生稳定且可解释的神经元集成动态。

本文为研究者使用网络构建了一个脉冲关联存储器，与传统体系结构比，它使用的脉冲更少，同时在高存储器负载下保持了较高的查全率。

42-406，利用脉冲神经网络和瞬态编码机制处理瞬态信号的运作机制。

这是一篇来自谷歌研究者研发 Ihmehimmdi 项目的论文。

该项目是研发如何利用脉冲神经网络和瞬态编码机制处理瞬态信号并探索其运作机制；研发全连接的脉冲神经网络对于瞬间编码的处理能力；通过在模型中使用突触转移函数测量模拟神经元系统在接收信号时膜电位的电势的升高及衰减过程，检查突触效率，并精确测定突触后脉冲时间对于突触前脉冲时间权重的导数。

论文摘要：对瞬态信息的保留有助于更好地表达像声音一类动态特征，同时可以对瞬态发生的时间实现快速响应。在生物中，尽管神经系统有几十亿神经元，但信息仅仅依赖于单个脉冲信号传递，单个神经元将信息编码到时间信号，并发出信号在神经网络中传输。信号脉冲本身包含了丰富的信息。

在神奇的生物特性启发下，Ihmehimmeli 项目聚焦于脉冲神经网络（SNN），利用不同架构和学习体制充分挖掘瞬态信号动力学行为。利用脉冲神经网络和瞬态编码机制处理信号，这样的编码机制天然存在于脉冲时序输入的神经元特征中，而

这种网络输出则由最早激活的输出神经元进行编码。

研究人员在最近发表的模型中阐述全连接脉冲神经网络对瞬间编码的处理能力。在模型中使用生物特性启发的突触转移函数，检查突触的效率，信号强度的变化是由连接的权重决定的，这代表突触的效率。

43-415，开发近红外光激发的纳米探针，监测大脑深层活动，理解神经系统功能机制。

开发、设计电压敏感纳米探针一直是个技术难关。

群体神经元活动的在体监测是揭示神经系统功能机制的关键。

近日《美国化学会志》期刊报导一项新的研究成果：研究人员开发了一种可用近红外光激发的电压荧光纳米探针，并用它监测斑马鱼和小鼠脑中神经元膜电位的动态变化。

目前神经元钙离子荧光成像是主要手段之一，但相比于神经脉冲信号，钙离子荧光信号的动力学相对较慢，且很难推断出与之对应的神经脉冲的频率和数量。因此神经科学界迫切期望能开发出对细胞膜电位变化敏感，有高信噪比的纳米粒子或分子探针，从而实现高时空分辨率、大范围神经元集群活动的活体监测。现有的荧光电压探针多用紫外光或可见光激发，只能应用于大脑浅层。而红外光(750-1000nm)在生物组织中穿透能力更强(可达 cm 级)，能应用于大脑深层，被称为“生物组织的光学窗口”。

研发高灵敏、可用于近红外光激发的电压敏感探针是目前国际神经科学领域迫切希望攻克的技术难关之一。

哺乳动物神经元膜电位的阈下振荡，反映动物个体的脑状态及其变化。

44-423, 来自 DeepMind、牛津大学和谷歌大脑的 Timothy P. Lillicrap、Adam Santoro、Geoffrey Hinton 在《Nature》子刊《Nature Reviews Neuro Science》发表文章，他们认为在以往研究的基础上，反向连接可能会引发神经元活动，而其中的差异可用于局部逼近误差信号，从而促进大脑深层网络的有效学习。

45-426, 新发现的大脑启发式加速学习机制（大大优于机器学习）。

发表在 Scientific Reports 杂志上的一篇文章中，研究人员报告说，他们已经重建了实验神经科学与高级人工智能学习算法之间的桥梁。在神经元文化上进行新型实验，研究人员能够证明一种新的、受大脑启发的学习机制。例如，当用于手写数字识别的人工任务时，其成功率大大超过了常用的机器学习算法。

研究人员证明了两个假设：关于大脑学习非常缓慢的普遍假设可能是错误的，并且大脑的动力学可能包括加速的学习机制。研究小组的实验表明，训练频率会大大加速大脑的适应。这种新发现的大脑启发式加速学习机制的使用大大优于常用的机器学习算法，例如手写数字识别，尤其是在提供小的数据集进行训练的情况下。

46-427, SNN 机理性测试。

SNN 利用时空处理，脉冲稀疏性和较高的内部神经元带宽来最大化神经形态计算的效率。尽管可以在这种情况下使用常规的基于硅的技术，但最终的神经元突触电路需要多个晶体管和复杂的布局，从而限制了集成密度。论文展示了双门高斯异质结晶体管的前所未有的静电控制，可简化尖峰神经元的实现。这些设备采用晶片级混合尺寸的范德华异质结，包括化学气相沉积的单层二硫化钼和溶液

处理的半导体单壁碳纳米管，以模拟生物神经元中产生脉冲的离子通道。基于这些双门高斯器件的电路可以实现多种异步脉冲响应，包括相位脉冲，延迟脉冲和强音爆发。除神经形态计算外，可调高斯响应对一系列其他应用（包括电信，计算机视觉和自然语言处理）也具有重要意义。

47-428， 脉冲神经网络 SNN 拟脑探索。

随着对生物大脑进行建模的计算机的需求不断增长，神经启发性计算机领域已发展到了脉冲神经网络（SNN）的探索，并克服了传统的冯-诺依曼架构所面临的挑战，基于硬件的神经形态芯片已被设计。神经形态芯片基于脉冲神经元，仅当它们接收到脉冲信号才处理输入信息。给定稀疏分布的输入脉冲序列，由于可以将不受传入脉冲驱动的网络的大部分设置为功率门控模式，因此可以减少此类事件驱动硬件的功耗。在硬件中构建具有大量突触的脉冲神经形态芯片需要解决的挑战包括构建具有低功耗的小型脉冲神经核，高效的神经编码方案和轻量级的片上学习算法。在这篇论文中介绍了用于 3D-NoCSNN 处理器的轻量级脉冲神经元处理内核（SNPC）的硬件实现和评估，以及其片上学习模块的设计。SNPC 嵌入 256 个 Leaky integrate and Fire (LIF) 神经元和基于交叉开关的突触，覆盖 0.12mm 平方的芯片面积。使用 MNIST 数据集对其性能进行评估，得出的推理精度为 97.55%。

48-432， Loihi—PohoikiSprings 大规模神经拟态计算系统。

今天，英特尔公布其迄今为止最大规模神经拟态计算系统 PohoikiSprings，包含 1 亿个神经元！这一大小，堪比小型哺乳动物的大脑容量。

英特尔神经拟态计算实验室主任 Mike Davies 表示：“PohoikiSprings 将我们的 Loihi 神经拟态研究芯片扩展了 750 倍以上，同时可在低于 500W 的功率水平下工作”。

如今，神经拟态计算已被普遍看作是解决人工智能等计算难题的重要途径。著名研究机构 Gartner 预测，到 2025 年，神经拟态芯片有望取代 GPU，成为先进人工智能部署的主要计算架构。

与传统处理器相比，英特尔 Loihi 芯片速度快 1000 倍，能效高 10000 倍。

英特尔今天公布的 PohoikiSprings 进一步扩展，将 768 颗 Loihi 神经拟态研究芯片集成在 5 台标准服务器大小的机箱中，形成了一个更强大的机架式数据中心系统。

这一系统将提供给英特尔神经拟态研究社区（INRC）的成员，以扩展其神经拟态工作来解决更大规模且复杂的问题。

49-433，英特尔神经拟态计算实验室主任 Mike Davies 接受芯东西等媒体线上采访。

提问：

从第一颗 Loihi 芯片诞生到如今发布最大神经拟态系统，英特尔神经拟态计算如何起步？

如今在各种计算架构中，英特尔神经拟态计算的技术特点是什么？它致力于解决哪些核心问题，又如何克服哪些技术挑战？

同样是支持大规模运算任务的前沿计算技术，神经拟态系统与量子系统有哪些不同？

随着 1 亿神经元的 Pohoiki Springs 系统问世，神经拟态计算发展到什么阶段？

英特尔是否已有明确的商业计划？

Mike 全面解读了英特尔神经拟态计算的研究进展、技术特性、应用示例与所面临的商业化挑战。

Mike 告诉大家，英特尔不打算做面向特定应用的专用芯片，而是希望能让神经拟态计算实现和冯-诺依曼架构相似的通用性。

Mike 说，距离将产品推向市场，还将等待数年。

50-434，英特尔 Mike Davies 解读神经拟态计算。

研发像大脑一样聪明的自主学习芯片，灵感来自交换机和摇头娃娃。

大脑神经网络通过脉冲传递信息，根据脉冲时间调节突触强度或突触连接的权重，并把这些变化存储在突触连接处。大脑神经网络与其环境中多区域间相互作用，就产生了种种智能行为。它能同时学习语音任务、视觉任务、决策任务、操作控制等许多任务，并且仅仅消耗非常低的能量。

在自然界，很多昆虫大脑的神经元远低于 100 万个，就能做到实时视觉跟踪物体、导航和躲避障碍物。

那么在硬件层面上能不能复制大脑神经元组织、通信和学习方式，打造出一种更智能的系统呢？这种颠覆传统计算机架构的新型计算模式就是神经拟态计算。

英特尔神经拟态计算的研究始于几十年前与加州理工学院 Carver Mead 教授的合作，他是半导体专家，今天我们需要芯片专业知识、物理学和生物学的结合，为这一创想提供了可行的土壤。Mike Davies 是英特尔神经拟态计算实验室主任，当年他在 Fulcrum Microsystems 研究以太网交换机，而交换机芯片异步设计适

合于神经拟态芯片的原型架构、算法、软件和系统的研究，它采用一种新颖的异步脉冲方式来计算。怎么让 Loihi 芯片模拟大脑进行学习呢？在芯片上运行简单的图像识别神经网络，先是看到物体，再输入物体名称，训练后就能快速准确辨认出是什么东西，摇头娃娃给研究者很大启发，使训练后的 Loihi 芯片能从人堆中识别特定对象。

神经拟态计算较之传统计算架构的优势在于超低延时、超低功耗。

在谈到与量子计算比较时他说，英特尔现在已经有了可使用支持 1 亿神经元的神经拟态系统，但距离拥有 1 亿量子比特的计算系统还很遥远。

英特尔于 2028 年 3 月建立神经拟态研究社区（INRC），使用 Loihi 芯片为研发活动的架构焦点，推动神经拟态算法、软件和应用程序的研发。

最后他说，神经拟态系统仍处于研究初级阶段，其设计目的并非取代传统的计算系统。

除英特尔外，IBM、惠普、麻省理工学院、普渡大学、斯坦福大学等顶尖研究机构，都在推进类脑计算的相关研究。

前路依然漫长，研究人员终极目标是接近当今性能最强大的计算机——人类大脑。

51-435，科学家在小鼠脑中发现攻击行为“调节开关”。

攻击行为是大多数动物重要的本能行为，也是躁狂症、双相情感障碍等精神疾病的表现形式。深入研究攻击行为发生的神经机制，可为病理性攻击行为提供新的治疗思路，最终为人类治疗病理性精神疾病服务。

由中科院神经科学研究所、上海脑科学与类脑研究中心、神经科学国家重点实验室许晓鸿研究员率领的研究团队对小鼠大脑皮层在攻击行为中的神经机制进行

研究，在小鼠大脑中发现攻击行为的“调节开关”。

小鼠感知外界入侵者的刺激后，从调整自身状态，到最终输出攻击行为，整个过程需要大脑内一系列“核团”的参与。其中，主要以“下丘脑腹内侧核”为核心，其他脑区与其形成直接与间接的连接，参与攻击行为的发生。“下丘脑腹内侧核”属于皮层下结构。

他们的研究论文近日在国际权威学术期刊《细胞报导》(Cell Reports)上在线发表。

52-438，神经拟态计算技术PK传统人工智能加速技术。

2017年，英特尔第一款自主学习神经拟态芯片Loihi问世（采用14nm工艺，每颗芯片125个内核，每个内核支持1000个神经元，单颗芯片支持13万个神经元、1.28亿个突触），接着开发PohoikiSprings（2020年3月发布），由768颗芯片集成组成神经网络支持1亿个神经元组网。从此Loihi发展成1亿神经元大规模的神经拟态计算系统。

神经拟态计算技术与传统的人工智能加速技术相比，有哪些技术特点？两者有哪些相同或不同？（包括所解决任务的不同）

神经拟态计算技术与人工智能加速技术两者都是为了处理人工神经网络而设计的，同样是支持大规模运算任务的前沿计算技术。

传统设计的AI芯片，即人工智能加速器是一类专用于人工智能（如人工神经网络、机器视觉、机器学习等）硬件加速的微处理器或计算机系统。据英特尔神经拟态计算实验室主任MikeDavies谈，英特尔不打算做面向特定应用的专用芯片，而是希望能让神经拟态计算实现通用性。

神经拟态计算与传统的人工智能加速器主要差异在哪里？

我们先从两者的定义出发：

众所周知，传统的AI芯片是一类专用于人工智能硬件加速的微处理器(CPU、GPU)或计算机系统。

所谓神经拟态计算技术是将人类大脑功能映射到硬件上，即直接用硬件（神经拟态芯片）来模拟人类大脑结构（这种方法叫神经拟态计算）。

两者最大区别在：

神经拟态技术研究类脑芯片严格按人脑的机制，将训练和推理整合到一块芯片（如Loihi）上，实现存储与计算融合；而基于传统冯-诺伊曼计算架构设计的AI芯片虽然声称是类脑芯片，但它并未严格按人脑机制，其训练与推理是分离的，存储与计算也是分离的。这样神经拟态计算技术变得更加智能、快速和高效，与传统处理器比，Loihi芯片速度快1000倍，能效高10000倍。

神经拟态计算技术相较传统架构的优势不在于吞吐量而在于超低延时和超低功耗。

国际著名IT调研分析公司Gartner认为：如今神经拟态计算已被普遍看作解决人工智能等计算难题的重要途径。预测到2025年，神经拟态芯片有望取代GPU成为先进人工智能部署的主要计算架构。

所谓Loihi芯片拥有并行性和异步信号传输能力，在这里谈到并行计算，公平地说两者是一样的，但神经拟态计算自然拥有异步信号传输能力的优势！它采用一种新颖的异步脉冲方式来计算，像人脑一样根据环境的反馈来自动学习如何操作，随着时间的推理，变得越来越智能、快速和高效。

以往国内外都在开发传统的专用的AI芯片（人工智能加速器），神经拟态系统仍

处于初级研究阶段，但除英特尔外，IBM、惠普、麻省理工学院、普渡大学、斯坦福大学等顶尖研究机构，目前均在紧跟开展这方面的研究。

53-442，人工智能与人脑愈来愈远还是愈来愈近？

神经网络之父 Hinton 的反思。1986 年 Geoffrey Hinton 写了一篇论文《Learning representations by back propagation errors》，将反向传播算法首次被引入到多层神经网络的训练中去，其重大意义更是为人工智能在最近十年的发展奠定了基础。

当下面对人工智能的局限性，面对人类大脑尚有很多未知的运作机制，Hinton 提出反思和质疑。

反思：我们是否需要放弃反向传播算法，重新开辟一条新路径？

质疑：人工智能与人类大脑愈走愈远还是愈来愈近？

国内 AI 专家黄铁军教授认为：人类大脑结构是亿万年“优胜劣汰”进化过程造就的。反向传播是人工智能的训练手段。

54-447，人工智能走向何方？

喜看三条出发路线，五位世界人工智能大师发评论。

人工智能与人脑越来越远还是近，也有大师发评论。

人工智能未来发展的第一目标是人类智能或接近人类智能。

目前有三条路，正在探索走向发展目标：

一、从深度神经网络（或机器学习/深度学习模型）出发

有人说，深度学习已近天花板，似乎很难往前发展了。他们说，深度学习是一个

强大的数据分析工具，带动了当前人工智能的繁荣，但它本质上也是一项暗箱技术或盲模型，其训练过程不可解释、不可理解、不可控，缺乏类人的推理能力，与人类大脑的运作机制差距很大，难以逾越；也有人说，深度神经网络潜力很大，自监督学习（训练）可使深度学习达到或接近人类智力水平，这时出现了发展的转机。说这话的人还是世界人工智能大师，如 Yoshua Bengio、Yann LeCun、Geoffrey Hinton 等，他们坦率地谈了未来深度学习人工智能的研究趋势，认为自我监督学习是一种机器学习/深度学习的“理想状态”，可使之产生类人的推动力，变不可解释、不可理解、不可控为可！去年人工智能算法大师 John-Hopcroft 更是信心满满要在 5 年内打破深度学习这个黑盒子，他说人类知道它在学习，但不知它怎么学习，我们会在 5 年内大体能读出深度学习的数学理论。

对于走这条路是否受限于天花板一直有争议，现在看来突破天花板发展有转机！

二、从异步脉冲神经网络出发

异步脉冲神经网络与人类大脑神经元网络在结构、特征、功能、机制等方面比较相似（或力求相似），因此它在对人类大脑意识处理的探索上比其他路径有优势，但我们对异步脉冲神经网络的研究还处于初级阶段，欲达到人类智能或接近人类智能的目标，还有很长的路要走，还会遇到很多挑战：

在神经形态计算出现后我们必须把传统的冯-诺伊曼计算架构转移到神经形态计算（类脑计算）架构上来，把目前采用的人工智能加速技术（AI 芯片）转移到神经网络拟态技术（芯片）上来（神经拟态芯片模拟人脑运作机制，主要采用异步脉冲神经网络）；我们应与神经科学联系，对异步脉冲神经网络很多未知的关键技术、运作机制和功能表现有待深入研究与工程实践：关键还要进一步深入理解人类大脑神经元的生物特性和运作机制以用于我们的研发；研发基于脉冲信号信

息处理的稀疏和时间的动态特性、脉冲时序编码机制、突触转移高效函数、异步脉冲传输机制及各项功效指标等；异步脉冲神经网络向类脑方向发展也离不开自监督学习和训练。

总之，对于走这条路很多人工智能专家是向往的，但日前研究尚处于初级阶段，迄今国内外均未拿出亮眼的成果。

三、从知识表示、驱动、推理，建设大规模语义网络出发

业内人工智能专家欢呼：2019 年自然语言处理（NLP）取得重大突破！

这条知识工程之路从感知智能奔向认知智能。上世纪 80 年代中期启动了知识工程，本世纪初又更新为新知识工程。新知识工程的重点是建设大规模语义网络（以提升知识图谱）。语义网络的发展过程是从自然语言处理系统到自然语言理解系统，再到大规模语义网络。IBM 沃森主张在以知识表示、驱动、推理的路上，由大规模语义网络支持的认知智能目标得以实现。

早年间，IBM “Watson Health” 搞医疗人工智能走的就是这条路。IBM 认为，对人工智能最重要的能力是知识而非数据。他们探索知识表示、驱动、推理，以期医疗人工智能从不可理解、不可解释的感知智能阶段推向可理解、可解释的认知智能阶段。但 IBM 走的这条路是失败的。

IBM 的失败，其中主要原因之一是大规模语义网络还不够完善，还没有能力支持认知智能的实现。这里我们引用图灵奖得主、人工智能大师 Yoshua Bengio 对此评论中的一段话：“NLP 虽然取得较大进步，但与人类相差还甚远”。

必须指出，对于常识、专业知识、专家经验，机器是很难识别的。IBM 提出具人（embodiment）概念，强调人工智能专家必须与临床医生结合，在疾病诊断时要取得共识。还有达到人类智能的另一道难题是：背景知识，这在学习和训练时是

不可或缺的。

所以对于走这条路，未来是非常有前途的，但当下还不成熟，路还很长。

55-464，改进脉冲神经网络的时间神经编码的新的训练方法——精确的突触效率调整方法。

脉冲神经网络（SNN）是第三代神经网络，在认知任务（例如模式识别）中表现出色。

在生物海马体中发现的时间神经编码机制使SNN拥有比具有其他编码方案的网络更强大的计算能力。但是这种时间编码方法要求神经元以离散化的方式处理信息，这大大降低了学习效率。为了保持时间偏码机制的强大计算能力，并克服其在SNN训练中的低效率，提出了一种新的训练方法，即精确的突触效率调整方法，受到灵长类动物视觉系统的选择性注意机制的启发，我们的算法仅选择目标脉冲时间作为关注区域，而忽略了非脉冲时间的电压状态，从而大大减少了训练时间。此外我们的算法采用了基于输出神经元电位与SNN触发阈值之间的电压差的代价函数，而不是传统的按照时间精确的时间间隔进行触发。

56-465，探索SNN节能硬件实现。

普渡大学研究生院探索生物类似的尖峰神经网络（SNN）的节能硬件实现。讨论SNN系统主要亮点是：①解决基于网络芯片（NOC）的SNN引起的连接性问题，②使用随机函数发生器（BRNG）提出随机CMOS二进制SNN。

57-466，移动智能设备/压缩传感技术/追踪去噪BPDN/稀疏图像编码器/LCA神经网络硬件实现。

当今的移动智能设备通常受到数据通信所需能量的限制，而不是数据处理所受的能量限制。因此，除了在信号处理中的传统用途外，压缩传感技术现在在低功率传感系统中的相关性越来越高。但是基本追踪去噪（BPDN）是此类技术所需的稀疏优化，通常计算量大，无法直接求解，因此实现通常采用近似于优化的追踪方法。

本地竞争算法（LCA）是一类脉冲循环神经网络，可解决 BPDN 加速器的高效硬件架构。这样的加速器是原型稀疏图像编码器，它使用定制的模拟神经元实现了无与伦比的能源效率。该工作已集成到数字设计流程中。混合信号原型的效率仅为 48.9pj/像素和 50.1nj/编码，是等效全数字架构的两倍。

当负责对手写数字图像进行编码时，该原型会生成稀疏代码，这些稀疏代码将压缩 90%以上，同时可以保留特征。与传统处理技术相比，原型压缩传感器雷达处理器将目标范围和速度估值的精度提高了 6 倍以上。该原型机能够每秒产生超过 100000 个估算值，与最新技术相比，吞吐量提高了 8 倍，效率提高了 18 倍。此外，由于突触权重压缩的独特形式，原型体系结构是迄今为止完全连接的 LCA 神经网络的最大硬件实现。

58-467，改进脉冲神经网络（SNN）训练方法在完成任务中获好性能并抗噪性。

近年来，基于梯度反向传播的脉冲神经网络（SNN）训练方法逐渐兴起。在这种训练方法下，SNN 能够在保留神经元内部动力学的同时获得较好的性能。

在此基础上，中科院自动化研究所听觉模型与认知计算团队，模仿刻画视听觉系统神经元侧向作用的数学模型动态神经场，提出了具有侧向作用的 SNN——LISNN 用于图像识别任务。并且在测试中，根据侧向作用的动力学特点，人为加入噪声

以验证侧向作用对网络鲁棒性的提升。

该研究工作分别在静态数据集 MNIST 与 FashionMNIST、动态数据集 N-MNIST 上对 LISNN 的性能进行了验证。输入数据以特定方式编码为一定长度的脉冲序列，每个时刻的序列规模与原图像（或事件点坐标范围）相同。模型在 MNIST 和 N-MNIST 数据集上均取得了和已有最好性能相近的结果；在 Fashion-MNIST 数据集上则取得 SNN 中的最好性能。

与传统的神经网络算法相比，所提算法取得较好的性能并自然地实现对噪声干扰的抗性，具有一定理论研究价值和工程实用价值。

59-487，优选脉冲神经网络（SNN）的不同训练方法以改善其特性（计算效率、准确性、推理延时）。

脉冲神经网络（SNN），通常被称为第三代神经网络，它使用稀疏脉冲事件在时间上对输入信息进行编码，可以利用它们来实现认知任务的更高计算效率。但考虑到最新的模拟神经网络（ANN）所带来的准确性的飞跃，SNN 训练算法的成熟度要低得多，从而导致了 SNN 和 ANN 之间的准确性差距。论文提出了不同的 SNN 训练方法，其生物保真度不相同，并评估了它们在复杂图像识别数据集上的功效。首先，论文针对 SNN 中无监督的表示学习提出基于生物学上合理的峰值定时依赖可塑性（STDP）的确定性和随机算法。论文中对 CIFAR-10 数据集的分析表明，基于 STDP 的学习规则使卷积层可以使用更少的训练示例来自学习底层输入特征。但是基于 STDP 的学习在适用浅层 SNN（ ≤ 4 层）时会受到论文提出了一种转换方法，以将现成的经过训练的 ANN 映射到 SNN，以实现节能推断。论文在 ImageNet 上展示了 VGG16-SNN 的 69.96% 的准确性。但是，ANN 到 SNN 的转换会导致较高的

推理延迟，以实现最佳精度。为了使推理延迟最小化，论文提出了基于峰值的误差反向传播算法，该算法使用脉冲神经元的可微近似。论文在 CIFAR-10 上进行的初步实验表明，与转换后的 SNN 相比，基于脉冲的错误反向传播可以有效地捕获时间统计信息，从而将推理延迟降低多达 8 倍。

60-489，用脉冲神经网络解决一类约束非 Lipschitz 的优化问题。本文针对一类约束非 Lipschitz 优化问题，提出了一个脉冲神经网络 (SNN)，其中目标函数是一个非光滑、非凸函数和一个非 Lipschitz 函数的和，其可行集为一个闭合凸集。使用平滑近似技术，通过微分方程，对提出的神经网络进行建模，可以轻松实现。在可行集中目标函数有界条件下，我们证明了可行集中任何初始点的 SNN 解的全局存在性和一致有界性。在平滑函数的 Lipschitz 属性下提供了 SNN 解决方案的唯一性。我们表明，SNN 解的任何累积点都是优化问题的固定点。数值结果包括图像复原，盲源分离，变量选择和最小化条件数，以说明理论结果并显示 SNN 的效率。

61-500，神经拟态计算 VS 深度学习。

三位 AI 专家（杜克大学陈怡然、浙江大学唐华锦、英特尔宋继强）谈基于脉冲神经网络的神经拟态计算与深度学习的关系

宋：神经拟态计算和深度学习的关系是兼容并蓄不是取代。对深度学习已有的擅长，如模拟人类视觉或自然语言交互的任务，让深度学习的网络去模拟。在其他方面，如英特尔的 Loihi 芯片做了嗅觉方面的工作，还有机器人操控，多模态甚至跨模态之间的知识存储，可用神经拟态计算去实现。

唐：在一些特殊场景中，如并不需要精确的计算结果，而需在一个实时环境中给出一个鲁棒响应时，神经拟态计算有绝对优势。

陈：目前两者实现的功能没有特别大的不同。神经拟态计算具有鲁棒性及实时性优势，这些只是在深度学习上提升而不是技术上突破。

宋：英特尔成立神经拟态研究社区(INRC)，这是开源开放的，要在应用方面突破，抓应用落地。

神经拟态计算未来前景的切入点：

①一个是非结构化数据、实时性要求高的场景，②多模态、实时的场景（如机器人、无人机、需要持续学习自适应学习的场景）。

62-503，忆阻性神经混合芯片加速推动人工智能技术发展

近来全球开展忆阻性神经混合芯片的研发十分火红。

来自俄罗斯的一则报导：俄罗斯罗巴切夫斯基州立大学与多国科学家合作，提出了一种“忆阻性神经混合芯片”（Memristive Neuro hybrid chip）的概念，芯片可用于紧凑型生物传感器和神经假体，该概念是基于神经细胞和微流体技术组合的现有和前瞻性解决方案，使“空间有序活动神经网络”的植入成为可能。

其实研发忆阻器技术在欧美早已开始了！忆阻器芯片具有非线性电阻记忆的独特性，在模拟信号处理系统具有广阔前景，此外还可用作电生理活动传感器，发挥信息积累和非易失性储存功能。

在清华大学自主研发的第二枚类脑芯片（发表在《自然》杂志上）也显示了忆阻器技术。今年2月27日由钱鹤、吴华强教授领衔的团队研发出一款基于多阵列忆阻器存算一体系统，在处理或训练卷积神经网络时能效优于GPU（以前沿用的

图形处理器芯片) 高两个数量级。

63-505, 神经调节算法——受人脑神经调节机制启发的人工智能新颖算法。

神经科学领域所发现的认知机理首次在多任务环境下发现了新的算法应用。

为利用人工智能更好地模拟人脑神经调节功能机制的关键打开了新视野。

尽管近年来人工智能领域取得巨大进步, 但离人类智能还很远。当前的人工智能技术允许训练计算机智能体可以专门有针对性地对它们进行人工训练, 以更好地执行某项特定任务。

人类通过使用一生中获得的技能, 能够非常有效地适应新情况。例如一个学会在客厅走路的孩子也会很快学会在花园里走路。因为所学习得到的走路技能与大脑突触的可塑性有关, 突触可塑性改变了神经元之间的联系。在客厅学习行走技能与在花园中行走所需快速适应技能则与神经调节过程相关。神经调节通过化学神经调节介质调整神经元本身的输入——输出特性:

模拟人脑神经突触的可塑性是当今人工智能所有最新进展的基础。

然而到目前为止, 还没有科学研究成果提出将神经调节机制引入到人工神经网络的方法。

为了攻克这个难题, 一个由人工智能专家和神经科学家组成的团队开发了一种基于人脑功能机制的新算法, 称为神经调节算法。这个新颖而卓越的算法可以创建能够执行训练期间未遇到的任务的智能体, 使智能主体能够自动适应未知情况。在神经科学中, 神经调节指一种神经传导过程。在此过程中, 一个特定神经元使用一个或多个神经传导物质来控制一系列神经元。被一小群神经元覆盖住的神经传递介质会在神经系统中大范围地被释放出来, 进而影响到许多的神经元。

64-510，减少训练人工智能巨大能源消耗。

MIT 减少训练人工智能神经网络巨大的能源消耗（释放 60 多万磅二氧化碳）将其降到 1/1300。

为了释放和提高人工智能算力，深度神经网络模型需要训练。大规模训练运行 AI 需要把海量的计算机服务器集中起来消耗巨大的电力来支持程序运转，这通常会排放大量的二氧化碳。

MIT 研究员 Amherst 研究发现（2019.6），训练运行一款自然语言处理 AI 模型可以释放超过 62.6 万磅二氧化碳。

MIT 研发团队在 4 月份发表的论文中提出一种训练运行神经网络的新方法，该方法在调试运行神经网络时释放的二氧化碳仅为当前所用方法释放的 1/1300。神经网络是模仿人脑功能的一套算法。

国际数据公司 2018 年度发布的白皮书预计，到 2025 年全球数据量将达到 175zb，这些数据都需要处理，而处理它们必然要耗费能源。

MIT 研发团队还提出 13 种利用 AI 帮助人类适应气候变化并减轻不利影响的方法，包括 AI 能帮助人类预测可再生能源的供求关系，规划大规模的碳排放项目。

65-517，神经科学打开了人工智能的黑盒子。

马格德堡大学（Otto von Guericke University）人工智能实验室 Sebastian Stober 团队研发在认知神经科学启发下的可解释人工智能技术的研究项目（CogXAI），即将运用认知神经科学的方法来分析人工神经网络，从而更好地理解它们的工作方式。

他们认为，人工神经网络 ANNs 是一种受自然大脑结构启发的自学习智能系统，

它就像生物神经系统一样，能够通过实例学习来独立解决复杂的问题。

在人类大脑中，这些网络是数以百万计的神经细胞通过化学信号和电信号相互交流组成的，而人工神经网络可以被理解为计算机程序，但由于其强大的学习能力和灵活性，近年来人工神经网络在“深度学习”下已成为智能系统开发的热门选择。

科学家通过大脑研究对人类大脑的学习行为有重要发现，而人工神经可以借鉴用来获得快速有效的学习行为，并研究 ANN 的自学习系统如何做出预测，和/或为什么会出错。

Stober 说，对自然大脑研究已经超过 50 年了，然而目前这种潜力很少被用于 AI 架构的开发。通过将神经科学的方法转移到人工神经网络的研究中，他们的学习过程也将变得更加透明和容易理解。通过这种方式，在学习过程的早期阶段就可以识别出人工神经元的故障，并在训练中加以纠正。

人工神经网络的正在迅速发展，通过使用高性能计算机，越来越多的人工神经元可以用于学习。然而这些网络日益复杂，甚至连专家也难以理解它们的内部流程和决策。如果我们希望未来能够安全使用人工智能就必须全面了解它的工作原理。

66-535- (二)，从异步脉冲神经网络出发

英特尔最近发布了神经拟态计算的最新进展，这是由异步脉冲神经网络出发，实行类脑计算的重大突破。

脉冲神经网络 (SNNs) 是神经拟态计算中一种全新的模型，以其 Loihi 芯片为基础的神经拟态计算系统 PohoikiSprings 包含 1 亿神经元。“神经拟态计算”或“类脑计算”指的是模拟人脑神经机制和运行方式有关计算。

神经拟态计算可以模仿人脑中自然神经元网络的方式将计算模块重新分布。颠覆传统冯-诺伊曼的硬件和软件架构，实现人脑的智能功能。神经拟态计算芯片体积小、功耗低、符合生物进化最本质优势。神经拟态计算在算法和芯片设计上可实现以低一千倍以下的功耗去完成同样效果的模型训练。

神经拟态计算的优势是功耗低、神经元的智能性和自主性（最大优势），不是单纯解决一个数据训练、模式识别问题，而是解决多模态感知、非结构化信息白的感知和推理。

神经拟态计算被认为引领下一代人工智能的主流计算模式。

神经拟态计算还发于发展的初始阶段，当前的重点是抓应用突破、应用落地（为此英特尔成立开源的神经拟态研究社区 INRC）。神经拟态计算未来前景的切入点：①非结构化数据、实时性要求高的场景，②多模态、实时的场景（如机器人、无人机和需要持续学习、自适应学习的场景）。

67-540，异步脉冲神经网络（SNNs）是模拟人脑神经元连接、信息处理和运行方式的模型，以实行神经拟态计算或类脑计算，这也是人工智能的一条升级之路（从弱人工智能跃升为强人工智能）。

在传统计算（非人工智能）领域，乃至在弱人工智能领域，是采用计算机运作机制、传统数据处理通用算法，主要依靠硅基芯片微处理器，遵循冯·诺伊曼计算架构；在强人工智能领域，是采用模仿人脑的运作机制、神经拟态或类脑算法，将模仿人类大脑功能的神经拟态计算技术映射到神经拟态芯片（硬件）上，遵循神经拟态计算架构（或新型类脑计算架构）；英国曼彻斯特大学研发的 SpiNNaker，是一种新颖的大规模神经拟态计算+并行计算体系，其灵感来自人脑的基本结构

和功能，因此也可归类于新型神经拟态计算或类脑计算架构，其中的并行处理可以认为，是为实现大规模神经网络的仿真提供高性能大规模并行处理平台；在人类智能（智慧）领域，则依据人脑生物学的运作机制进行计算。

神经拟态计算或类脑计算是下一代人工智能的重要方向（或主流模式）。

当前计算技术发展遇到瓶颈：一是摩尔定律将失效，神经拟态芯片将取代传统的硅基微处理器芯片；二是冯·诺伊曼计算架构（读取、存储、处理数据能效低下）将被新的计算架构（类脑计算架构或神经拟态计算架构）所取代。

在进行学习、认知、处理等复杂计算时，人脑计算功耗最少，约 3-5W，类脑计算功耗约 350-500W，而标准计算机能耗约 50-200KW。据 Loihi 神经拟态计算研发团队报告：神经拟态计算比传统计算，速度快 1000 倍，能效高 10000 倍。神经拟态计算依靠神经元、突触进行信息编码和处理，神经元（计算）、突触（存储），计算存储不再分离，配置一体化，传输依靠脉冲（尖峰）神经网络，当尖峰运作时才进行信息传输，如此的传输机制有利于节能。

基于异步脉冲神经网络的神经拟态计算与传统计算运行方式比较表

	信息源	编码方式	传递方式	基于不同特征的计算方式
传统计算运行方式	数字信号 计算机程序	由数字源代码变换 0,1 的机器码进行编码	传统网络权重连接+激活方式，对机器码进行信息传递	传统计算机运行方式，符合冯·诺依曼硬件和软件计算架构
神经拟态计算运行方式	神经电脉冲信号和化学信号	采用稀疏脉冲时序编码机制	通过由仿脑自然神经元+突触组成的脉冲神经网络和运行方式模型，对神经电脉冲进行信息传递	神经拟态计算系统运行方式，打破冯·诺依曼硬件和软件计算架构

人脑或类脑计算模式与硅基计算机原理之间差别很大：

- ① 人脑或类脑计算中，计算（神经元）和存储（突触）是一体化的，融合在一起的，训练和推理也是一体化的，在硅基计算中，计算（处理单元）和存储（存

储单元) 是分离的, 训练和推论也是分离的;

②在人脑或类脑计算中, 存在三维广泛连通性, 在硅基技术上尚无法模拟三维连通, 受限于二维连接;

③在人脑或类脑计算中, 是基于脉冲的事件驱动型的随机计算, 在硅基技术中主要为了构建确定性计算采用晶体管间布尔代数电路开关。

④在人脑中是生物学自主低能耗类型, 在类脑计算中实行节能型, 在标准计算机中能耗大。

⑤在人脑或类脑计算中, 将模仿大脑的大规模并行通信方式(或结构), 将数十亿信息同时发送到数千个不同的目的地, 不会像传统通信模式那样通过标准网络从A点到B点发送大量信息来进行通信。

目前国内外研发异步脉冲神经网络与神经拟态计算(或类脑计算)系统案例如下:

① 英特尔案例

2017年发布 Loihi 脉冲神经网络(SNN)芯片。

芯片制程 14nm, 管芯尺寸 60mm, 每颗芯片包含 20 多亿个晶体管。

每颗芯片 128 个核, 每个核支持 1000 个神经元, 每颗芯片支持 13 万个神经元、1.3 亿个突触。

2019年, 英特尔将 Loihi 芯片集成为 PohoikiBeach:

64XLoihi=800 万个神经元, 组成神经拟态网络。

2020年3月19日, 英特尔将 Loihi 芯片集成为 PohoikiSprings: 768XLoihi=1 亿个神经元, 组成神经拟态网络(相当于小型哺乳动物大脑)。

应用场景

英特尔以开放姿态成立 INRC(神经拟态研究开源社区)以期在应用方面实行突破,

做好应用落地。

目前其神经拟态计算切入点为：

- 1) 基于神经拟态计算开展超算应用开发，
- 2) 针对非结构化数据、实时性要求高的场景进行应用开发，
- 3) 针对多模拟、实时的场景进行引用开发

（如机器人、无人机等需持续学习或自适应学习的场景）

② 浙江大学案例

似乎是模拟英特尔的案例

2015 年 12 月发表达尔文—1 脉冲神经网络芯片。

2019 年发表达尔文—2 脉冲神经网络芯片：每颗芯片 576 个核，每个核支持 256 个神经元，每颗芯片支持 15 万个神经元。

2020 年 9 月 1 日，浙大将达尔文—2 集成为：

792X15 万=1.2 亿个神经元（相当于小鼠大脑）。

应用场景正在起步

③ 国曼彻斯特大学案例

由曼彻斯特大学、南安普敦大学、剑桥大学、谢菲尔德大学、ARM、Silistix 公司、泰雷兹等共同研发，2016 年发布 SpiNNaker—1 脉冲神经网络（芯片）和类脑计算机，每颗芯片支持 1.6 万个神经元，曼彻斯特大学将 SpiNNaker 芯片集成为：30KX16K=4.6 亿神经元（SpiNNaker 的模拟规模相当于一只老鼠的大脑或十分之一个人脑。

68-541，异步脉冲神经网络与神经拟态计算系统

神经拟态计算技术与非人工智能的传统计算技术（及具有人工智能加速技术的传统计算系统）的区别在哪里？

有人说：“目前的人工智能本质上还是计算机科学的一个分支（现在国际上将人工智能的论文都统计在计算机科学的前沿研究中）”，对这种模糊的说法，似有写这篇补充的必要。

在这里我们还是引用《评人工智能如何走向新阶段？》（国内外跟贴留言）第四集（386-486 条）中的 438 条来自国外的留言：

神经拟态计算技术（新一代 AI）PK 传统人工智能（深度学习 AI）加速技术

神经拟态计算技术与传统的人工智能加速技术相比，有哪些技术特点？两者有哪些相同或不同（包括所解决任务的不同）？

神经拟态计算技术与人工智能加速技术两者都是为了处理人工神经网络而设计的，同样是支持大规模运算任务的前沿计算技术。

传统设计的 AI 芯片，即人工智能加速器是一类专用于人工智能（如人工智能网络、机器视觉、机器学习等）硬件加速的微处理器或计算机系统。据英特尔神经拟态计算实验室主任 Mike Davies 谈，英特尔不打算做面向特定应用的专用芯片，而是希望能让神经拟态计算实现通用性。

神经拟态计算与传统人工智能加速器主要差异在哪里？

先从两者定义出发：

传统的 AI 芯片是一类专用于人工智能硬件加速的微处理器 (CPU、GPU) 或计算机系统。所谓神经拟态计算技术是将人类大脑功能映射到硬件上，即直接用硬件（神经拟态芯片）来模拟人类大脑结构（这种方法叫神经拟态计算）。

两者最大区别在：

神经拟态技术研究类脑芯片严格按人脑的机制，将训练和推理整合到一块芯片上，实现存储与计算融合，而基于冯-诺依曼计算架构设计的 AI 芯片，它并未严格按人脑机制，其训练与推理是分离的，存储与计算也是分离的。这样神经拟态计算技术变得更加智能、快速和高效，与传统处理器比速度快、能效高，神经拟态计算相较于传统架构的优势不在于吞吐量而在于超低延时和超低功耗。

所谓 Loihi 芯片拥有并行性和异步信号传输能力，在这里谈到并行计算，公平地说两者是一样的，但神经拟态计算自然拥有异步信号传输能力的优势！它像人脑一样，根据环境的反馈来自动学习如何操作，随时间推理变得越来越智能、快速和高效。

人类大脑神经网络通过脉冲传递信息，根据脉冲时间调节突触强度或突触连接的权重，并把这些变化存储在突触连接处。在硬件层面上能不能复制大脑神经元组织、通信和学习方式，打造出一种更加智能的系统呢？这种颠覆传统计算架构的新型计算模式就是神经拟态计算。

以往国内外都在开发传统的专用 AI 芯片（人工智能加速器），近来神经拟态系统的研究在国外开始火红起来，英特尔、IBM、惠普 MIT、普渡大学、斯坦福大学、浙江大学等研究机构，均在紧跟开展这方面研究。

69-542，神经拟态计算，探索智能互联时代的计算创新

英特尔中国研究院院长 宋继强

（在《第十五届开源中国开源世界高峰论坛》线上会议的报告）

大家上午好，非常高兴来到这次开源中国开源世界的高峰论坛，跟大家分享英特尔在神经拟态计算方面的进展。我今天的报告主题是神经拟态计算探索智能互联时代的智能创新。

随着无线通信技术和人工智能技术的广泛应用，现在信息世界里的数据呈现了三大主要变化，首先是数据量，有指数级的爆发，最开始都是人工产生的数据，人往系统里输入数据，后来有了多媒体的数据。现在这个阶段，我们发现能够产生数据捕捉数据的设备越来越多。除了人使用的设备之外，还有大量物联网设备产生的数据，还有人工智能的很多模型自己也有很多数据，包括参数和训练的结果，这些称为原数据。

这些数据量上升很快，形态跟原来的结构化数据有很大的不同，结构化数据指的是人为设定的数据结构，比较易于管理和存储，但是非结构化数据通常是来自很多自然界的的数据，有很多处理的变化。

同时，很多数据刚才讲的来自于物联网设备的，来自于前面实际应用场景的，需要在终端或者说边缘侧被采集，同时有实时处理再反馈到边缘侧的需求，也就是说我们对于数据实时处理的需求也在快速增长。

应对这样的变化，我们觉得未来的十年，处理这些数据需要的计算架构，就不会是单一的计算架构，而是需要好多种不同计算架构去混合处理。架构创新将会成为计算创新的关键驱动力，目前来看主流的计算架构有这么四种，图上表示了四种主流计算架构是标量的、矢量、矩阵和空间。标量指的一个一个去计算数据，有前后的顺序。矢量就是把一组数一起来做，就像把两个数组直接相加相乘。矩阵把一块数据同时做运算。空间说明这些数据本身有很强的稀疏性，它的处理需要采用特殊的方法加速。现在的人工智能领域的数据处理，至少是矢量级别的，很多是矩阵的，在某些应用场景需要空间数据结构来加速。

这些整合在一起，称为 XPU 超异构计算，X 代表有多种可能性。超异构是什么概念？传统异构领域，要么把不同计算的能力，不同计算架构整合到一个单芯片上去，那是传统的 SOC 的做法。或者把不同的芯片，比如说 CPU、GPU、IPGA 整合到一个板上，板级的整合，这是另外一种异构计算。这两种各有优劣势。

超异构计算希望把他们的优势集中起来，而把不足之处去掉。也就是说我们可以把已经做好的很多种不同架构的加速的芯片单元通过异构封装的方式组合到一个芯片里面，把高性能的芯片之间的传输通道建立起来，这样就避免了在板级异构的方式，牺牲了空间也牺牲了传输的效率，同时又可以把很多不同架构芯片整合在一起，小空间、低功耗的情况下提高计算的性能和能效比。

这个图是直接的例子，就是我们怎么样把多种不同的处理器放到一个大的芯片封装里面，利用上高速芯片间的 IO 还有封装领域的高级技术，这是非常重要的技术趋势。

继续往前看我们觉得架构方面需要更多的探索，现在常用的两种一种叫标准计算的思路，我们已经知道一个问题，同时我们也能够对这个问题进行很好的建模，我们可以用一个流程图的方式把怎么处理它的步骤写出来，根据它去编程，这是由人工选定怎么做，人工实现怎么把它实施到软件或硬件中去，我们称为标准计算模式。

第二种模式是现在深度神经网络给我们带来的，也就是说我们知道输入的数据是什么，我们也知道想产生的结果是什么，但是中间的计算模型我们不清楚怎么做，人工很难去描述它。比如说人脸识别，到底是怎么算的模型，不允许我们描述。但是我们有大量的数据，可以端到端的进行训练，有神经网络、深度神经网络模型可以作为框架进行训练，最后训练出一个符合特定任务解决的方式。这种是叫数据驱动的计算的架构和模式。

第三种未来要更多考虑怎么把现在已经有的优势整合起来，因为标准计算来自于人类已经有的知识，来自于一些推理。而深度神经网络利用的是数据本身内涵的规律，计算的性能要求很高，对内存的使用要求很高，相应的功耗也会很高。如何能够达成优势互补，同时我们在最后使用的时候，又是能源友好，因为我们知道人类大脑只需要 20 瓦就能完成很多不同种类的事情，既可以做标准的逻辑推理运算，也可以做深度神经网络擅长的视觉识别、语音识别的运算。

我们要把它整合到一个架构里去，希望能够有一个能效比高的方案，同时又能够支持扩展。神经拟态计算是这方面的可行之道，我们来介绍一下它的进展。Loihi 神经拟态芯片，用了整合计算和存储架构来做的，不像传统计算一样，有 CPU，有内存，中间有数据传输通道，把计算单元和需要的存储空间紧密的整合在一起，形成非常多的小的核。

一个芯片里面就有 128 个小的核，每个核都是一样的，叫同构多核，每个核里可以做 1000 个神经元的模型，这个神经元就是在模拟人脑神经元的结构和数据处理和存储方式。每一个神经元又可以和另外的 1000 个神经元互相连接，也就是说构造人脑里面类似突触这样的概念。这样一个小的芯片就可以模拟人脑的 13 万个神经元，并且可以产生 1.3 亿个突触，已经是一个比较复杂的神经网络的突破。

它体积很小，能耗很低，需要使用的时候，只有工作那部分电路才耗电，其他电路可以保持非常低功耗的休眠状态，它是一个真正的绿色芯片，当世界越来越多的使用人工智能的时候，我们会需要这样高能效比，有自己的学习和省电能力的芯片。

我们也认为这是一种助力 AI 从 2.0 时代向 3.0 时代迈进的，为什么？因为 AI 的 2.0 时代更多利用数据驱动，做端到端的学习，去学习对识别，对大规模计算领域，数据量比较大领域的好用的模型。AI3.0 时代，我们希望 AI 能力可解释，可

以通过少量的数据去自己学习，不一定需要非常大量的数据，因为很多领域大数据是不一定容易获得的。同时，又希望它能够比较鲁棒，对恶意输入的数据，有抵抗能力。同时希望把多种不同的学习能力整合在一起，也就是说这么多种能力，希望在 3.0 时代获得。

Loihi 因为灵活的架构，可以支持多种事件驱动的学习模式。比如常规使用的无监督学习模式，这是在 2.0 时代深度学习已经非常常用的模式，一种是给出目标的具体的标签，另外一种是不用给出标签，主要做分类。

右边这两种是在未来各种各样的自主系统里面非常有用的，一种叫自监督学习，是关联不同的事件，去产生出新后的因果顺序，并且找出可能的关联，并且强化它。强化学习应用在自动驾驶、机器人领域或者交互式推荐这些方面，它是利用的观察到事件和作出的决策之间的关系，去反复做迭代，优化出一个最好的决策。这么多种不同的学习模式，在 Loihi 上面都可以编程支持，并且同时去学习。

Loihi 系统从 2017 年推出第一个单芯片，后面逐渐扩大它的规模，大家看这条线，从这几年来来看，我们已经能够用 768 个芯片集成在一个大的系统里面，这样可以提供接近 1 亿个神经元的规模，这样的规模已经接近于小的仓鼠哺乳动物的脑。有这样的规模未必现在有这个智能，这还要靠很多软件、很多算法去达到这样的能力，但是我们已经可以提供这样的规模，就是硬件的基础。

同时，这件事情本身又不是英特尔一家可以完全搞定的，还是需要广大的研究社区，还有很多企业参与，去寻找最好的应用领域和未来很多组合式创新的方法。英特尔成立了一个研究社区，是把大家聚合起来，一起来去推动这个领域的发展，这个研究社区是面向全球的。

我们现在可以提供计算的平台和能力，不光是提供在设备上直接的使用能力还提供在云端使用这些服务和测试系统的能力。举一个使用的案例，这是低功耗使用的案例，最近我们也发表了一个合作，用一片 Loihi 芯片训练，让 Loihi 系统拥

有了嗅觉。可以对十种不同的气味做识别，而且通过每个气味单一样本来训练，就可以让识别量达到 92%。传统深度学习的方法，至少要用 3000 倍以上的训练样本，才能达到这样的识别效率。同时，功耗也很低，Loihi 用的功耗非常低。

对于崭新的硬件架构，如何对它编程进行很好的能力释放是非常重要的，在结合之前我们讲的 GPU、硬件加速器，如果把它们统一做成一个异构系统，如何能让开发者或者说科研人员更方便的去使用它，做更多的创新，因为我们说未来不可能有人把所有的底层硬件都了解的很清楚。

对于英特尔来讲，这件事情非常重要，异构计算需要统一可控的编程模型，也需要软硬结合的方法去创新。我们针对不同硬件底层会提供直接的开发的性能库，包括底层的编程接口，但是在上层，对开发者来讲提供统一的 API，把这些能力释放出来，让在不同硬件之间去迁移能力变得非常简单和高效。这块就是我们称为通过 oneAPI 软硬融合，释放异构计算潜能的概念。希望大家来参与这方面的交流和合作，因为 oneAPI 是面向开源领域的。

最后总结一下，英特尔是通过创造改变世界的技术来造福地球上的每一个人，我们非常希望通过异构计算，通过开源的技术和平台，能够让新的技术快速覆盖到全球，产生造福人类的技术。谢谢大家。

70-543，2018 年 11 月 SpinNaker 类脑计算机首次亮相，为实时大规模神经网络仿真提供高性能处理平台。

这是一种新颖的大规模神经拟态计算+并行计算的体系结构，打破冯-诺伊曼计算架构和常规超级计算规则，它采用适合于大型尖峰神经网络建模（或基于脉冲神经网络）进行信息传送，SpinNaker 可实时模拟大量生物神经元，其规模相当于一只老鼠的大脑，或千分之一个人脑。

它不同于传统计算机，它是一台由百万处理器驱动的计算机，更像人类大脑，采用神经拟态+平行处理以与人脑类似的方式传递数据，它不会通过标准网络从 A 点到 B 点发送大量信息来进行通信，相反，它模仿大脑的大规模神经拟态+并行通信体系结构，同时将数十亿的信息同时发送到数千个不同目的地，实现构建用于实时的大脑建模应用程序的 100 万个核心，利用该机器，研究人员将能够模拟 10 亿个简单的神经元或数百万个具有复杂结构和内部动力学的神经元。

SpinNaker 项目启动于 2006 年，由曼彻斯特大学、南安普敦大学、剑桥大学、谢菲尔德大学及 ARM、Silistix、泰雷兹等公司合作研发。

类脑计算机主要用于处理其擅长的人工智能任务，其神经拟态计算架构打破冯-诺伊曼计算架构，但不会替代冯-诺伊曼架构的传统计算机，两者互补、融合可能会是未来的趋势。总的来说，在国际上类脑计算发展还处于初级阶段，未来走向成熟需要硬件、软件和算法的进步。

71-544，2020 年 9 月 1 日浙江大学联合之江实验室发布：成功研制全球神经元规模最大的类脑计算机 Darwin Mouse。

该类脑计算机含 792 颗达尔文—2 代类脑芯片（2019 年 8 月发布，15 万个神经元，相当于果蝇），该类脑计算机在对芯片集成处理后具有 1.2 亿个脉冲神经元和近千亿个神经突触，堪比小鼠大脑神经元规模，典型运行功耗仅 350-500 瓦，其运行机制与 SpinNaker 相同，但浙大声称完全是基于自主知识产权类脑芯片的类脑计算机。类脑计算机与传统计算机差别非常大，传统计算机的算法和软件无法在类脑计算机上运行。类脑计算机主要用于处理人工智能任务。

72-545, 2020年10月14日清华大学计算机系张悠慧团队、精仪系施路平团队合作在《Nature》杂志上发文,提出类脑计算完备性(类脑计算即神经拟态计算)以及软硬件去耦合的类脑算法层次结构(首次提出)。

与通用计算机的“图灵完备性”概念与“冯诺依曼”体系结构相对应,本文题为《一种类脑计算系统层次结构》(A system hierarchy for brain-inspired computing)的论文,通过理论论证与原型实验证明,该类系统的硬件完备性与编译可行性,并扩展了类脑计算系统使之能支持通用计算。类脑计算处于起步阶段,国际上尚未形成公认的技术标准与方案,这一成果填补了完备性理论与相应系统层次结构方面的空白,利于自主掌握新型计算机系统核心技术。

近年来类脑计算研究受到了越来越多的关注。类脑计算是借鉴生物神经系统信息处理模式和结构的计算理论、体系结构、芯片设计以及应用模型与算法的总称:在未来10-20年内,谁要引领世界经济,谁就必须在这个领域领先。

现在类脑计算系统方面的研究多聚焦于具体芯片、工具链、应用和算法的创新实现,而对系统基础性问题,例如计算完备性、系统层次结构等思考不足,导致软硬件紧耦合、应用范围不明确等一系列问题。但从现有通用计算机的发展历史与设计方法论来看,完善的计算完备性与软硬件去耦合的层次结构是计算系统蓬勃发展的计算理论与系统结构基础。

图灵完备性和冯诺依曼体系结构是通用计算机技术能够飞速发展并持续繁荣的关键因素——几乎所有的高级编程语言都是图灵完备的,冯诺伊曼架构通用处理器则可以通过图灵完备的指令集实现图灵完备性,这意味着编程语言编写的任何程序都可以转换为任意图灵完备处理器上的等价指令序列(即程序编译)。这样,由软件层、编译层、硬件层组成的计算机层次结构就能够确保应用软件、指令集、硬件设计在独立发展的同时相互兼容(即软硬件去耦合),为整个领域的繁苇发展打下了系统基础。

“完备性”可以回答系统能够完成什么、功能边界在哪里等问题。研究完备性，可以为软硬件系统的解耦合、划分不同研究领域的任务分工与接口提供理论基础。

73-554，类脑芯片（或神经拟态芯片）几例：

①IBM True North 芯片，2014 研发、2017 发布。

28nm 制程工艺，芯片尺寸 100 mm^2 ，每颗芯片 4096 个内核，每个内核支持 100 万个神经元、2.5 亿个突触每颗芯片支持 41 亿个神经元、1 万亿个突触。

芯片集成：48 颗芯片，由芯片集成组成的神经网络：2000 亿个神经元组网相当于普通老鼠大脑。

②英特尔 Loihi 芯片，2017 发布。

14nm 制程工艺，芯片尺寸 60 mm^2 。

每颗芯片支持 125 个内核，每个内核支持 1000 个神经元、25 万个突触。

每颗芯片支持 12.5 万个神经元、1000 万个突触。

由芯片集成组成的神经网络：1 亿个神经元组网，相当于小型哺乳动物大脑
Loihi 芯片训练后能效提高 1000 倍。

③清华大学 天机芯（TianJic），TianJic-1 2017 发布。

TianJic-2 2019.7.3 在 Nature 上发表

TianJic-3 尚在研制中

28nm 制程工艺，芯片尺寸 14.4 mm^2 。

每颗芯片 156 个内核，每个内核支持 4 万个神经元、1000 多万个突触。

每颗芯片支持 625 万个神经元。

由单芯片组成的神经网络（625 万个神经元网络）。

将基于脉冲神经网络（SNN）的类脑算法和基于人工神经网络（ANN）的深度学习算法集成到一颗芯片上。

④ 浙江大学+之江实验室， 达尔文-1 芯片， 2015 年发布

达尔文-2 芯片， 2019 年发布

达尔文-3 芯片， 尚在研制中

每颗芯片 576 个内核， 每个内核支持 256 个神经元， 6 万个突触。

每颗芯片支持 15 万个神经元。

达尔文-2 芯片集成： 792 颗芯片支持 1.2 亿个脉冲神经元、 300 亿个突触， 相当于果蝇大脑神经元数量规模。

⑤其他单位类脑芯片

曼彻斯特大学： SpinNaker 2018.11.2 发布。

其芯片集成： 支持 100 万个内核、 10 亿神经元， 相当于小鼠大脑神经元规模。

74-555， 类脑计算(神经拟态计算) 与传统硅基计算、 传递、 运行模式

①前者的信息源是神经电脉冲信号和化学信号， 后者为数字信号

②前者编码方式采用稀疏脉冲时序编码机制， 后者由数字源代码变换为 0,1 的机器码

③前者信息传递方式通过模仿人脑自然神经元+突触组成的脉冲神经网络和运行方式模型， 对神经电脉冲进行信息传递， 后者采用传统网络权重连接+激活方式， 对机器码进行信息传递

④前者计算（神经元）和存储（突触）是一体化的、融合在一起的，后者计算（处

理单元)和存储(存储单元)是分离的

⑤前者存在三维广泛连通性,后者无法模拟三维连通,受限于二位连接

⑥前者是基于脉冲的事件驱动型的随机计算,后者为了构建确定性计算采用晶体管间布尔代数电路开关

⑦前者类脑是模拟生物学上自主低能耗类型(计算/处理、传输),实行节能型,后者能耗大

⑧前者模仿人脑的大规模进行通信方式(结构),将数十亿信息同时送到数千个不同目的地,后者也可采用并行通信方式,但传送信息为通过标准网络(从A点到B点),发送大量信息进行通信

⑨前者其运行方式符合类脑计算(神经拟态计算)系统的运行方式(或神经拟态计算架构),后者其运行方式符合冯·诺依曼计算架构。

75-556, 类脑计算完备性

这是清华大学计算机系张悠慧教授团队和精仪系施路平教授团队合作撰写的一篇文章(论文作者为:Youhui Zhang, Peng Qu, Yu Ji, Weihao Zhang, Guangrong Gao, Guanrui Wang, Sen Song, Guoqi Li, Wenguang Chen, Weimin Zheng, Feng Chen, Jing Pei, Rong Zhao, Mingguo Zhao & Luping Shi)。论文题目:

《A system hierarchy for brain-inspired computing》,是清华大学今年第三次在《Nature》杂志上发表的文章(2020.10.14):

本文摘要:提出神经拟态完整性

大脑启发计算目前缺乏一个简单而健全的系统层次来支持整体开发。因此,神经

拟态软件和硬件之间没有清晰完整的接口。由于许多灵感来自大脑的芯片不是为传统的通用计算而设计的，它们中很少提供传统的指令集，因此不清楚它们是否是图灵完成的。而图灵完备性是传统编译的可行性基础，要求程序的表达和转换是等价的。所以本文提出神经拟态完备性，这是一种更适用于大脑启发计算的完备性的更广泛定义。它放宽了神经拟态硬件的完备性要求，提高了不同硬件和软件设计之间的兼容性，并通过引入一个新的维度——近似粒度来扩大设计空间。

提出一种系统层次结构

神经拟态计算与传统计算的区别还在于：它使用同步计算和存储，使用基于spikes (spiking 神经网络的特征) 的事件驱动计算，并在高并行方面具有更大的潜力。这些差异使得传统的计算机层次结构难以直观地描述类脑应用程序。因此提出了一种具有高通用性的类脑计算系统层次结构。这个层次结构有三个层次：软件、硬件和编译。

软件层次，提出了一个统一的、通用的软件抽象模型——编程操作符图 (programming operator graph, POG)——以适应各种大脑启发的算法和模型设计。该模型集成了存储和处理。它描述了什么是大脑激发程序，并定义了它是如何执行的。由于POG是图灵完成的，它最大程度地支持各种应用程序、编程语言和框架。

硬件层次，设计了抽象神经拟态体系结构(ANA)。

编译层次，是将程序转换为硬件支持的等效形式的中间层。为了实现可行性，提出了一套被主流的脑激发芯片广泛支持的基本硬件执行原语，证明了配备这套硬件的神经形态是完整的。最终通过实验验证了神经形态完备性引入的系统设计层次的优化效果。

相关实验部分

第一个应用实验是一种用于自行车驾驶和跟踪的人工神经网络模型。

第二个应用实验是用于鸟群模拟的 boids 模型。

第三个应用实验是 QR 分解（非线性计算的数学算法）。

神经拟态计算从生物大脑中汲取灵感，为计算技术和体系结构提供了推动下一波计算机工程发展的潜力。这种受大脑启发的计算也为人工智能的发展提供了有前途的平台。传统的计算机系统具有围绕图灵完备和冯·诺伊曼体系结构建立的完善的计算机层次结构，而与传统的计算机系统不同，目前尚无广义的系统层次结构或对类脑计算的完整性的理解。这会影响软件和硬件之间的兼容性，从而阻碍类脑计算的开发效率。

清华大学该团队提出了“类脑计算完备性”，它放宽了对硬件完整性的要求，并提出了相应的系统层次结构，其中包括图灵完备的软件抽象模型和通用的抽象神经形态架构。使用这种层次结构，可以将各种程序描述为统一的表示形式，并转换为任何神经形态完整硬件上的等效可执行文件。也就是说，它可以确保编程语言的可移植性，硬件完整性和编译可行性。该团队实现了一系列工具链软件用以支持在各种典型的硬件平台上执行不同类型的程序，进而证明了系统层次结构的优势。

希望可以使类脑计算系统的各个方面实现高效且兼容的进展，从而促进包括人工智能在内的各种应用程序的开发。

76-562，清华、北大教授同台激辩：脑科学是否真能启发人工智能？

教授们认为，脑科学是智能科学一个重要的研究方向，但不是做人工智能的前提，

它与人工智能应该是一个相辅相成的过程。

目前认知神经科学取得极大进步，对现在的人工智能的端对端学习与强化有诸多启发，但认知科学的进展对下一代人工智能是否有帮助？脑科学是否能真正启发人工智能？都是有待探索的问题。

有的教授提出，能否构造不同于人脑的认知智能系统？

进化不是线性的，是一棵进化树。不同生物有不同的神经元，但所有生物（高级的、低级的）对世界都有非常好的适应。所以无论其神经元多少，都能做到通用智能。

不同生物的神突触连接是不一样的，据此如果完全按照生物系统以物理上的神经结构去做智能，真的能够得到普遍性的规律吗？

人工智能有很多途径：主流的是由数理逻辑和专家系统去做符号主义；另外一个连接主义，构造一个大的神经网络；还有行为主义，主要是从控制论去做。人脑是一个近乎完美的通用智能系统，所以当连接主义出现问题和瓶颈时，人脑的存在可提供一个最终的方向和信念。

计算机里的任何一个问题，并没有最优算法，都是多个最优算法同时存在，所以实现人工智能也应该有多条路径。

如果我们想从人类智能或生物智能借鉴一些东西帮助人工智能发展，最核心的是理解人脑，这需要去深入挖掘。

现在的人工智能是通过计算机、机器实现的，其根本的是要适应机器而不是人脑。对脑科学或人工智能最重要的应该是存在性的东西，智能是存在的，但是脑科学目前并没有提供一种途径实现人工智能。

下一代人工智能最需要去关注什么？

哪些人类或生物体的认知功能是下一代人工智能最需要去关注或借鉴的？

当下生物体智能和人工智能有哪些差异？

大脑与人工智能在结构上不同，结构不同实际是编码集不同。

神经系统还有 Conference（来自先验知识），通过神经元反应的随机性把 Conference 编码在一起，机器学习中的 Conference 是通过大数据做出来的，现在的问题是，神经系统发生的机制和机器学习是否一样？

在有限资源下重要的是学习、记忆及遗忘。知识是人区别于动物最本质的东西。

脑神经是比较简单的物理系统，把它变成向量在逻辑上是通的，但是向量比较精确，脑能否处理这么复杂的事？

现在的知识图谱是三元组，三元组做推理是没有问题，但脑子里是不是三元组？

大脑的神经元是多样性的，但目前神经网络里的神经元都长一个样子，去做这样一个网络模型很困难。

大脑是复杂的网络动力学模型，引入神经元里的动力学系统后，能做记忆和联想，如何把这种类似于神经动力学的系统反映在一个计算系统上是一个比较关键的问题。

大脑里有很多可以借鉴的东西，如稀疏编码、注意力机制，能把多层次多精度的记忆和联想在计算系统里做实现。而现有计算机体系存算分离，计算系统中所有记忆都一样。如神经网络存的全是权重，没有把记忆分层次粒度。因此如何把记忆跟计算系统及决策联系起来也是一个关键问题。

人工智能如何与各学科联合发展？

77-563, 创建结构数学研发通用 AI

大脑是模拟的，机器（计算机）是数字的，深度学习将数字输入大脑，大脑是不识别的，而大脑中的思维逻辑传到机器，机器也是不识别的。所以采用深度学习算法技术难以做到脑机交互。

研发下一代通用人工智能，从计算理论角度讲，巴贝奇与图灵计算模式均不适用，需要创建新的计算理论基础，统一时空计算理论，并与脑科学统一；目前的障碍在于数学描述方法，要创建一门新的数学——结构数学，以解决结构化多因果描述全息结构计算模型。

78-564, 设计 2—16—256 进制类脑计算机

现在已经实现设计 16 进制的类脑计算机。这时不仅计算速度、计算能力能呈指数级增长，关键是计算性能更符合人脑功能机制。

79-567, 构建“类脑智能”（或“仿脑”）不是一完全“克隆人脑”（或“复制类人”）有人提出下一阶段人工智能即通用人工智能（强人工智能），其目标是构建“类脑智能”，不应该是构建一个与人脑完全一样的东西，是一个模拟人脑的智能，不是完全克隆人脑的“类人智能”！

80-576, 神经拟态计算为何被认为是下一代 AI?

——杜克大学陈怡然教授、浙江大学唐华锦教授、英特尔中国研究院宋继强院长
热议神经拟态计算

神经拟态计算日益“火热”!

英特尔最近公布了在神经拟态计算的最新进展。英特尔 2017 年开发的异步脉冲神经网络 Loihi 芯片 (SNN), 仅需单一样本便可学会识别 10 种有害气体的气味。以 Loihi 芯片为基础的神经拟态计算系统 PohoikiSprings 包含 1 亿个神经元, 堪比小型哺乳动物的大脑容量。

神经拟态计算或类脑计算或神经形态计算, 指的是机器模拟人脑神经机制和运行方式的有关计算。脉冲神经网络 (SNNs) 是神经拟态计算中一种全新的模型, 可以模仿人脑中自然神经元网络的方式将计算模块重新分布。颠覆传统冯-诺伊曼的硬件加软件架构, 实现人脑的智能功能, 神经拟态计算被认为是引领下一代人工智能的主流计算模式。

三位学术界大咖在这次行业对话中着重对神经拟态计算的独特优势, 与目前主流的深度学习的对比, 如何进一步突破, 以及未来应用方向进行交流。

解决现有 AI 的挑战。目前人工智能正面临诸多挑战, 其中一项就是对于能源的大量需求, 造成对生态的污染问题。

仅仅训练一个 AI 模型, 消耗的电力 (产生的碳排放量) 相当于 5 台美式轿车整个生命周期的碳排放量。可以说目前的 AI 模式对生态环境构成了一定的威胁。

作为下一代 AI 芯片, 神经拟态计算能很好解决这一问题。

宋继强: 神经拟态计算在算法以及芯片的设计上可实现以低一千倍以下的功耗去完成同样效果的模型训练。

唐华锦: 芯片体积小、功耗低, 符合生物进化最本质优势。

神经拟态计算优势：功耗低、神经元的智能性和自主性（最大优势）不是单纯解决一个数据训练、模式识别问题，而是解决多模态感知、非结构化信息的感知和推理。

陈怡然：神经拟态计算比目前技术更加安全。可以通过不同信号相互间连接做得更“鲁棒（Robust）”，这对于外在的攻击或不友好操作可更有效进行保护。

神经拟态计算与深度学习的关系。

神经拟态计算是否在不远将来取代深度学习？

Gartner 调查报告预测：到 2025 年神经拟态计算有望取代 GPU，成为下一代 AI 主流计算形态。

宋认为，神经拟态计算和深度学习的关系是兼容并蓄，不是取代。对深度学习已擅长模拟人类视觉或自然语言交互的任务，还是让深度学习的网络去模拟。在其他方面，如 Loihi 芯片做的嗅觉方面，还有机器人操控、多模态甚至跨模态之间的知识存储，可用神经拟态计算去实现。

唐认为，在一些特殊场景中，如并不需要太精确的计算结果，而需在一个实时环境中给出一个鲁棒响应时，神经拟态计算有绝对优势。

陈认为，目前两者实现的功能没有特别大的不同。神经拟态计算具有鲁棒性及实时性优势，这些只是在深度学习上提升而不是技术上突破。

英特尔现在成立神经拟态研究社区（INRC）期望在应用方面有所突破，抓应用场景落地。

英特尔已研制出 PohoikiSprings(1 亿神经元)神经拟态计算机样机(非诺架构)并投入试用。与传统计算机比，运算速度提高 1000 倍，能耗降低 10000 倍。

神经拟态计算未来前景切入点：

- ①一个是非结构化数据，实时性要求高的场景，
- ②多模态的、实时的场景（如机器人、无人机），
- ④ 要持续学习、自适应学习的场景。

81-587，英特尔公布神经拟态新进展，探讨神经拟态计算的商业应用，更新 Loihi 神经拟态研究基准

——浙江大学融媒体中心谢子怡报导，2020.12.5

在 12 月 4 日英特尔研究院开放日上，英特尔神经拟态计算应用合作伙伴：联想、埃森哲、罗技、梅赛德斯-奔驰、机器视觉传感器公司 PropheSee 等与英特尔一起分享英特尔神经拟态研究社区（INRC）的最新应用成果（进展）及其在商业应用上的价值。这次英特尔重点介绍 Loihi 神经拟态研究基准，包括：

■ 语音命令识别

埃森哲测试了英特尔 Loihi 和标准 GPU 上识别语音命令的能力，发现 Loihi 不仅达到了和 GPU 类似的精度，而且能效提高 1000 倍以上，响应速度快 200 毫秒。梅塞德斯-奔驰也正在探索如何将这些结果应用到现实中，如在汽车中加入新的语音交互命令。

■ 手势识别

埃森哲展示了 Loihi 在快速学习、识别个性化手势方面取得的切实进展。只需几次曝光，Loihi 即可学习新手势，可用于智能产品交互，或者公共场所非接触显示。

■ 图像检索

零售行业的研究人员评估了 Loihi 对基于图像的产品搜索应用，发现在保持相同精度水平的情况下，Loihi 生成图像特征向量的效率，比传统 CPU、GPU 方案提升

3 倍多。英特尔此前研究发现, Loihi 在百万幅图像数据库中搜索特征向量的速度比 CPU 快 24 倍, 且能耗低 30 倍。

■ 优化和搜索

英特尔发现 Loihi 解决优化和搜索问题的效率比传统 CPU 提高 1000 倍, 速度快 100 倍, 此研究可用于无人机实时规划并做出复杂导航决策, 也可以扩展到复杂的数据中心负载, 完成协助列车调度、物流优化等任务。

■ 机器人技术

罗格斯大学和代尔夫特理工大学的研究人员展示了在 Loihi 上运行机器人导航、微型无人机控制应用。代尔夫特理工大学的无人机使用一个包含 35 个神经元且能演进的脉冲网络进行光流着陆 (Opticflowlanding), 频率超过 250HZ。罗格斯大学发现, 在同等性能下, Loihi 解决方案的功耗比传统 GPU 低 75 倍。Loihi 还可以成功学习诸多 OpenAIGym 的任务, 精度与深度行动者网络 (DeepActorNetwork) 旗鼓相当, 而能耗比移动 GPU 解决方案低 140 倍。英特尔还展示了 Loihi 如何自适应地控制水平跟踪无人机平台, 实现最高 20KHZ 的闭环速度、200 微秒的视觉处理延迟比传统方案提高 1000 倍。

82-588, 用于训练脉冲神经网络的多代理进化机器人框架

普渡大学 SouvikDas, AnirudhShankar, VaneetAggarwal

2020.12 发表

自然界中广泛存在着竞争进化的机制。受此启发, 作者提出一种基于多代理进化机器人 (ER) 的新框架用于训练脉冲神经网络。脉冲神经单元的权重和它们控制的机器人的形态参数被视为表型, 框架规则根据某些机器人和它们的脉冲神经单元在生存和繁殖过程没有明确的变异, 但它们通过狩猎食物以及在规则中存活来

隐式驱动进化。它们捕获食物的效率在训练中表现出点缀平衡的进化特征。作者演示了表型上的两种进化继承算法——突变和突变的交叉，比单纯突变有着显著优势，提升了 40% 的学习能力。

83-596，迈向可持续的千倍速计算未来

——英特尔中国研究院院长宋继强

回顾 2020 年，新冠肺炎疫情给我们的工作和生活模式带来了重大变化。其中，有一项技术正在潜移默化地改变我们的生活，即 5G。2020 年是 5G 商用的关键之年，中国的 5G 技术部署在这一年取得了突飞猛进的进展。5G 是由技术驱动的创新，早在约 10 年之前，5G 技术就开始投入研发。之所以要大力拓展 5G 技术，并不是为了迎合当时的需求，而是看到了未来对于带宽和网速的需求潜力，是为“未来”做出的技术布局。

5G 技术的进展让我想到了近期英特尔研究院开放日的活动主题——“追求计算的千倍提升”。类似于 5G，要满足未来的计算需求，即超高带宽、超低时延、超大规模连接的需求，我们需要一种“超前”思维。因此产业现在就要开始提前布局，追求计算的千倍提升，在目前智能化、数字化的大背景下，这种“超前”思维非常有必要。

以超前思维布局未来计算范式

数字化、智能化已经成为不可阻挡的趋势，受到今年新冠肺炎疫情的影响，这一趋势以更快的速度席卷而来。如今，全球已有超过 100 亿台设备与云中的超级计算机实现了互联，未来这一数字将增长到 1000 亿。拥抱数字化不是选择题，而是必选题。在全民数字化的浪潮之下，数据量呈爆发式增长态势，数据形式也更

加多元化，可以说，未来的计算需求将有千倍速的提升。英特尔追求计算的千倍提升，就是从计算的供给侧出发，为未来的计算需求构建坚实基础。

除了“超前”思维之外，要想实现计算的千倍提升，还需要“超常”思维，即要打破常规。随着数据越来越多元化，新的计算范式不再是锦上添花，而是雪中送炭。常规和传统的单一架构已经不能满足越来越复杂的计算需求，未来需要更快、更灵活、更低功耗的“新计算”来破题。

软硬件双突破释放千倍算力

这种“超常”思维将在以下几个领域得到体现。首先，在硬件方面，需要打破单一架构，多架构融合的 XPU 架构将成为主流。XPU 架构不仅能大幅提升算力，同时还能够根据需求进行快速组合，降低成本，灵活性高。英特尔是目前全球唯一一家已经覆盖这四种主流芯片的厂商，得益于先进的封装技术，英特尔正在异构计算领域突飞猛进。

除此之外，面向未来，还需要对架构本身践行“超常”思维。举例来说，颠覆传统的冯·诺伊曼架构，模仿人脑神经元结构的神经拟态计算芯片就是一个很好的例证。这种芯片的优势在于可以在提升性能的同时大幅降低能耗。英特尔及其合作伙伴发现，英特尔神经拟态计算芯片 Loihi 解决优化和搜索问题的能效比传统 CPU 高 1000 倍，速度快 100 倍，已经实现计算速度的千倍提升。

其次，在软件方面。XPU 架构的诞生，给软件提出了更高的要求，因为能够同时掌握多种架构编程语言的开发人员凤毛麟角，而软件是释放硬件性能的关键一环，因此能够跨架构编程的软件模型以及可以提升编程效率的工具就显得极为重要。为此，英特尔也提前布局，跨架构编程的统一模型 oneAPI Glod 版本已在本月正式发布，将在很大程度上解决跨架构编程的难题。

未来算力要强大也要绿色

要实现算力的千倍提升，还需要坚持可持续发展的原则。千倍速的提升不能以千倍的功耗为代价，可持续发展是实现千倍提升的必要条件。

目前，计算对于能源的需求巨大。有研究报告显示，训练一个大型 AI 模型所产生的碳排放量相当于 5 辆美式轿车整个生命周期所消耗的碳排放量。因此面向未来计算的千倍提升，只有坚持可持续发展原则，才是真正符合人类利益的技术进步。

英特尔在技术发展方面一直坚持可持续原则。已经有结果显示，作为下一代 AI 芯片，英特尔神经拟态计算芯片 Loihi 在处理语音命令识别时，不仅达到了和 GPU 类似的精度，并且能效提高 1000 倍以上。除此之外，英特尔最新的集成光电技术将光子技术与硅芯片紧密集成，可以最大限度地缩小硅光子设备的体积，从而降低成本，将对数据中心进行彻底革新。

诸如此类的例子在英特尔还有很多，英特尔的宏旨是“创造改变世界的科技，造福地球上的每一个人”，通过我们的“超前”思维、“超常”思维以及可持续发展的原则，英特尔正引领产业迈向千倍速的计算未来。对这一天的到来，我充满期待。

84-604, 在记忆性脉冲神经网络中学习 STDP 模式识别的必要条件

Demin V. A. 等, 莫斯科大学和莫斯科国家研究中心

发表于《natural network》2020.12

本文目的在于研究实验和理论方法，通过高性能的基于记忆器的脉冲神经网络 (SNN) 寻找有效的本地训练规则，实现无监督模式识别。首先，使用脉冲——定

时依赖性 (STDP) 进行权重变化的可能性研究, 通过 CoFeB 连接的一对硬件模拟神经元来演示纳米合成记忆器。接下来, 在单层完全连接的馈送 SNN 的多种记忆 STDP 参数中, 对二元聚类任务解决方案的学习收敛进行了分析。显示出神经元之间具有优胜者——接受——全部竞争。为了研究训练收敛的基本条件, 作者构建了一个基于速率的单层网络的原始概率生成模型, 该模型具有独立或相互竞争的神经元, 并进行了深入分析。主要结果是“相关性增长——反相关衰减”原则, 它说明要接近最佳策略来配置模型参数。可定义为最佳学习的必要条件。最后, 本文对一种启发式算法进行了实验, 找出记忆 SNN 中的收敛条件。该方法具有通用性, 在搜索本地规则以确保其在模式识别任务域中的无监督学习收敛时, 可应用于软件或硬件速率编码单层 SNN 的多种电子学和神经元。

85-605, 使用高成功率和高精度的量子加速脉冲神经网络

——北京大学、中国邮电、日本 KDDI 研究所

本文提出了一种量子脉冲神经网络, 对其进行了综合评估, 并给出了详细的数学证明, 包括成功概率、计算精度和算法复杂度。证明表明量子 SNN 的计算复杂度在数据维度上是对数多项式。此外, 本文还提供了一种将量子 SNN 的最小成功概率提高到接近 100% 的方法。最后本文展示了量子 SNN 在解决现实世界中模式识别方面的良好性能。

86-606, 脑电与计算机视觉的交叉

——普渡大学电子计算机工程学院，2020.12.23 发表

发表在人工智能顶刊（TPAMI）上的一篇文章表示脑电与计算机视觉的交叉存在缺陷和盲点。

利用 EEG 测量技术，对观看 ImageNet 刺激的受试者的大脑诱发进行分类，并采用从该处理中获得的表示来构建一种新型的目标分类器是当下热门的研究方向。研究结果说明了存在于所有神经影像数据中的时间自相关对分类实验有深远影响。

87-628，Disout 算法 PK Dropout 算法

看到跟帖 626，谈到 Dropout，这是一个谷歌闭源的算法（谷歌已申请专利），谷歌定义 Dropout 为“解决神经网络过拟合的系统和方法”。

这使我想起跟帖 481：“华为开源 Disout 算法 PK 谷歌闭源 Dropout 算法”，重录如下，为跟帖 626 的补充。

在跟帖 481 中谈到：华为的 Disout 是一种通过研究特征图扰动来增强深度神经网络的泛化能力的方法。Disout 算法超过 Dropout 算法，如在 ImageNet 上训练的 ResNet—50，可达到 78.76% 准确率，而 Dropout 训练仅为 76.8%。

88-658，神经形态（拟态或类脑）芯片算法之优化

北京大学微纳电子研究院黄如院士在第 66 届国际电子器件大会（IEDM）上作神经形态器件报告

北大类脑智能芯片研究中心蔡一茂教授、黄如院士课题组为神经形态计算的器件一阵列一算法协同优化设计提供指导

阻变器件是后摩尔时代构建新型存算一体及类脑芯片、突破冯-诺依曼体系结构瓶颈的关键电子器件技术之一。但阻变器件的非理想效应以及高密度集成带来的热效应会相互耦合，成为阻变器件在存储及神经形态计算应用中的关键挑战。

蔡一茂教授、黄如院士课题组，研究阻变器件非理想效应的物理机制，提出了准确描述多种非理想效应的集约模型，建立了能够综合评估器件技术、阵列拓扑及算法设计的跨层次验证平台，掌握了非理想效应和热串扰对存储及神经形态计算应用的影响，为器件-阵列-算法的协同优化设计提供了重要指导。

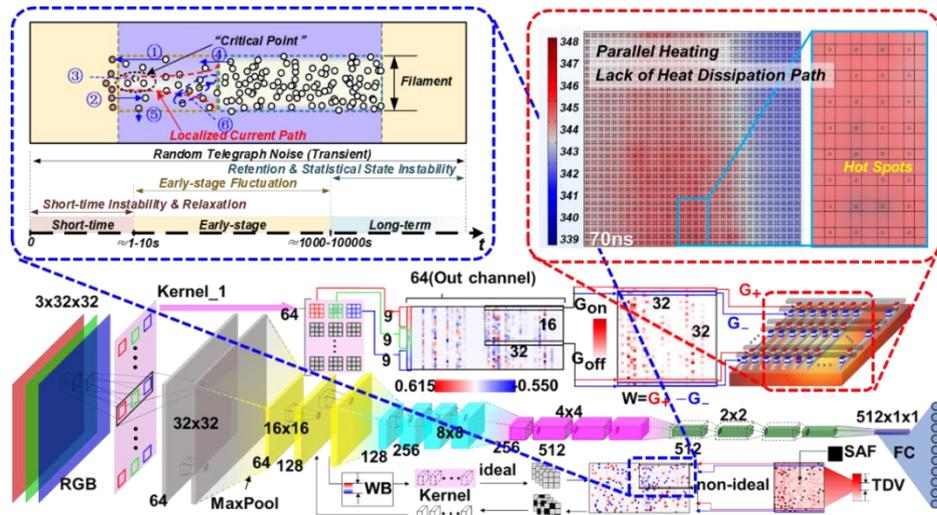
89-660，阻变器件——存算一体的类脑芯片

北京大学黄如院士研发团队

北大微纳电子研究院黄如院士在第 66 届国际电子器件大会（IEDM）上（2020.12.12-18）上发表他们研发的神经形态器件（类脑智能芯片）的论文。

阻变器件是后摩尔时代构建新型存算一体及类脑芯片、突破冯-诺依曼体系结构瓶颈的关键电子器件技术之一。但阻变器件的非理想效应以及高密度集成带来的热效应会相互耦合，成为阻变器件在存储及神经形态计算应用 1 中的关键挑战。

蔡一茂教授、黄如院士团队系统研究了阻变器件非理想效应的物理机制，提出了准确描述多种非理想效应的集约模型，建立了能够综合评估器件技术、阵列拓扑及算法设计的跨层次验证平台，掌握了非理想效应和热串扰对存储及神经形态计算应用的影响，为器件-阵列-算法的协同优化设计提供了重要指导。



90-662, 带有自适应脉冲递归神经网络的准确高效时域分类

Bojian Yin, Federico Corradi, Sander M. Bohte

荷兰阿姆斯特丹大学

(荷兰阿姆斯特丹大学在 arxiv 平台上发表一篇将脉冲神经网络应用到时域相关问题的文章)

受更详细的生物神经元建模启发, 脉冲神经网络 (SNN) 已被研究为生物学上更合理, 潜在能力更强大的神经计算模型; 但是, 与传统的人工神经网络 (ANN) 相比, 此类网络的性能仍然不足。在这里, 我们在时域中具有挑战性的基准问题 (语音和手势识别) 领域, 提出了 SNN 的最新技术—将最新的替代梯度与可调谐和自适应脉冲神经元的递归网络结合。这一技术超出了标准经典递归神经网络 (RNN) 的性能, 并且接近最佳现代 ANN 的性能。由于这些 SNN 表现出稀疏的脉冲, 因此我们证明, 与具有可比性能的 RNN 相比, 它们在理论上的计算效率高出 1-3 个数量级。总之, 这可以将 SNN 定位为 AI 硬件实现的有吸引力的解决方案。

91-667, 论系统软件在神经拟态计算能量管理中的作用

Twisha Titirsha, Shihao Song, Adarsha Balaji, Anup Das

德雷克塞尔大学, 2021. 3. 22

最近, 诸如 DYNAP 和 Loihi 的神经拟态计算系统被引入计算机界, 以提高机器学习程序的性能和能效, 尤其是那些使用 Spiking Neural Network (SNN) 实现的机器学习程序。用于神经拟态系统的系统软件的作用是将大型机器学习模型 (例如具有许多神经元和突触) 聚类并将这些聚类映射到硬件的计算资源。本文考虑了神经元和突触所消耗的功率, 以及在互连上传递尖峰时所消耗的能量, 从而制定了神经拟态硬件的能量消耗。

基于这样的表述, 本文首先评估系统软件在管理神经拟态系统能量消耗中的作用。采用一种简单的基于启发式的映射方法, 将神经元和突触放置到计算资源上以减少能耗。本文还用 10 个机器学习应用程序评估了本文提出的方法, 并证明了所提出的映射方法可显著减少神经拟态计算系统的能耗。

92-668, 使用脉冲间隔对脉冲神经网络进行视觉解释

Youngeun Kim, Priyadarshini Panda

耶鲁大学, 2021. 3. 26

脉冲神经网络 (SNN) 可以与异步二进制时间事件进行计算并进行通信, 这可以通过使用神经形态硬件来大大节省能源, 最近 SNN 相关算法工作已显示出在各种分类任务上的良好性能。然而, 目前缺乏研究用于分析和解释这种深度 SNN 的内部脉冲行为的可视化工具。在本文中, 我们提出了一种针对 SNN 的生物可视化的新概念, 称为峰值激活图 (SAM)。拟议中的 SAM 通过消除计算梯度以获得视觉解释

的需要，规避了脉冲神经元的不可微特征，相反，SAM 通过在不同时间步长上正向传播输入脉冲来计算时间可视化图。SAM 通过突出显示具有短峰间间隔活动的神经元来产生与输入数据的每个时间步相对应的注意力图。有趣的是，在沒有反向传播过程和类标签的情况下，SAM 会突出显示图像的区分区域，同时捕获细粒度的细节。借助 SAM，我们首次根据优化类型，泄漏行为以及何时遇到对抗性示例，对内部脉冲在各种 SNN 训练配置中的工作方式进行了全面分析。

93-676，利用神经拟态网络预测癫痫的全新算法

Fengshi Tian, Jie Yang, Shiqi Zhao, Mohamad Sawan

CenBRAIN 实验室

现有的较为有效的癫痫发作预测方法多为卷积神经网络 (CNN) 算法，他们都具有特异性和敏感性。但是 CNN 在计算上很昂贵并且耗电。这些不便之处使基于 CNN 的方法难以在可穿戴设备上实现。本文受高能效尖峰神经网络 (SNN) 的启发，在这项工作中提出了一种全新的用于预测癫痫发作的神经拟态算法。这种方法使用设计的高斯随机离散编码器从 EEG 样本生成尖峰序列，并在结合了 CNN 和 SNN 优势的尖峰卷积神经网络 (Spiking-CNN) 中进行预测。实验结果表明，与 CNN 比，灵敏度、特异性和 AUC 分别可得持 95.1%、99.2% 和 0.912，计算复杂度降低了 98.58%，表明所提出的 Spiking-CNN 具有硬件友好性和高精度。

94-681，基于异构网络和混合编码的快速智能神经拟态传感器

Angel Yanguas-Gil

阿贡国家实验室 2021.4.9

本文以昆虫的大脑为模型，了解如何利用异构结构（结合不同类型的神经元和编

码) 来创建集成输入处理、评估和响应的系统。本文还展示了如何利用时间和速率编码的组合构建高速传感器, 该传感器能够在几个周期内基于输入生成假设, 然后将该假设作用辅助输入以进行更详细的分析。

95-687, 脉冲神经网络中无监督模式识别的权重发散促进原理

Oleg Nikitin, Alex Kunin, Olga L. Ukyanova

俄罗斯科学院, 2021. 4

信号处理任务与生物神经元之间的并行性加深了人们对输入信号识别自组织优化原理的理解。本文讨论了生物学和技术系统之间的相似之处, 提出了对著名的 STDP 突触可塑性规则的补充, 以将权重调整指向与背景噪声和相关信号之间的最大差异相关的状态。物理上限制权重增长的原理被用作这种控制权重改变的基础。有人提出, 生物突触的直接修饰受到可塑性发展所需的生物化学“物质”的存在和生产的限制。在本文中, 有关信噪比的信息用于控制此类物质的产生和存储, 并驱动神经元的突触压向具有最佳信噪比的状态。本文进行了使用不同输入信号体制的几个实验, 以了解所提出方法的功能。

96-692, 基于近似误差反向传播的脉冲神经网络

Matteo Cartiglia, Germain Haessig, Giacomo Indiveri

瑞士苏黎世大学神经信息学研究所, 苏黎世理工学院

2021. 4. 23 发表

脉冲神经网络 (SNN) 在低功耗传感处理和边缘计算硬件平台的设计方面展示了广阔的前景。然而, 在这样的架构上实现片上学习算法仍然是一个开放性的挑战,

特别是对依赖于反向传播算法的多层神经网络而言。在本文中，我们提出了一种基于 SNN 的学习方法，该方法使用局部权重更新机制进行近似的误差反向传播操作，可以与混合信号模拟 / 数字神经形态电路兼容。我们设计了一种网络结构，该体系结构使突触权重更新机制能对误差进行跨层的反向传播，并提出了一种网络，该网络可以通过训练来区分具有相同平均放电率的两种模式。这项工作迈向了具有片上学习能力电路的超低功率混合信号神经形态处理系统设计的第一步，该电路可以通过训练来对不同的时空模式进行识别（例如基于事件的视觉或听觉产生的时空模式）。

97-694，比较类脑计算和传统计算的不同运行方式

	类脑计算运行方式	传统计算运行方式
信息源	神经电脉冲信号和化学信号	数字信号（或模拟信号）
编码方式	采用稀疏脉冲时序编码机制	由数字源代码变换为 0, 1 机器码进行编码
计算模式	计算（神经元）和存储（突触）是一体化的，融合一起 存在三维泛连通性，基于脉冲的事件驱动型的随机计算	计算（处理单元）和存储单元是分离的，无法模拟三维连通，受限于二维连接，构建确定性计算
传递方式	通过由仿脑自然神经元+突触组成的脉冲神经网络和运行方式模型，对神经电脉冲进行信息传递	传统网络权重连接+激活方式，对机器码进行信息传递
基于不同特征的计算方式	类脑计算（拟态计算）系统的运行方式，打破冯-诺伊曼计算架构	传统计算机的运行方式符合冯-诺伊曼硬件和软件计算架构 而且一、类脑算法，低能耗，二、传统算法，能耗大。

98-696，人工智能如何在非冯架构上表现？

问：目前所有的人工智能的成就都是在计算机上表现出来，是基于冯架构的计算机智能或计算智能，人工智能是计算机的一个应用而已。而人脑不是冯-诺依曼架构的，存在不存在宏观上更类似脑的非冯-诺依曼架构呢？例如，对人的智能

而言，记忆力是真正的智力，超强记忆力就是超强智能，记忆比计算机重要，记忆的提取要比复杂的推理快得多，非冯架构如何在结构上体现人脑的不同记忆区和记忆力呢？如何体现环境和知识的双驱动？

答：存在非冯计算结构是肯定的，基于准全息系统论与结构计算模型，我们已设计实现 16 进制类脑计算机，在此基础上设计 256 进制计算机也完全没有问题！这是典型的非冯结构计算机。不仅能体现储算一体化特征，更能体现环境和背景知识双驱动的本质特征。

人工智能迄今已产生若干理论和技术流派，但用硬件实现的成果较少，显现智能的突破性进展有限。迄今为止，按生物神经网络（BNN）巨量并行分布方式构造的各种人工神经网络，并未体现人们所期盼的智能。专家系统在一个阶段的成应用后逐步显露其局限性。历时 10 年耗资 540 亿日元的第五代计算机，在研制可实用化的计算机方面并没有完成预期目标。在涉及现场问题处理及实时获取知识方面遇到瓶颈，过多的规则和预先规定限制了系统的柔性和适应性，如模式识别能力较差，增强处理问题的适应性成了首要难题。在事实驱动和目标引导的双向推理方面未能引入实时交互作用机制。神经网络、知识工程、行为机制和现场学派都是从不同侧面模拟局部人脑智能，不能从系统构成的理论层面提供有益启示。人脑智能模拟必须基于自然系统理论才能解决。软件代替不了符合自然原理的硬件。寻求自然计算原理的突破是构造类脑计算机的唯一正确选择，试图用软件模拟人脑智能此路不通！智能是多层次、多子系统的功能互补效应，不能在一个功能层次完整体现。

设计类脑计算机既解决计算机与背景信息相融统一问题，又解决计算机与人脑近似同构同功问题。

类脑计算或智能模拟必须解决三大问题：一，解决状态之间的自组织模式（存储单元的内在联系）问题，二，解决计算原理及效率（与人脑同构同功）问题，三，解决多功能层次的统一运行机制及双向交互作用（功能耦合）问题。同时解决计算机与环境及背景信息分离的问题，解决传感、效应子系统与核心信息处理机制分离问题，解决储算分离问题，解决总线瓶颈、二值逻辑、确定性及形式化悖论等问题，这些问题绝非传统理论及技术能够解决的。软件本质上是发挥硬件功能，硬件不具备的功能软件同样无能为力，因而不能指望软件模拟人脑智能。

计算机存储单元之间没有内在联系，子系统功能耦合与状态的自组织机制分离，状态转换与环境作用机制分离。而人类状态关系结构与状态记忆、处理结构是统一的。在转换状态时，状态与环境及背景信息统一，是智能本体（自然）状态关系的自组织。基于这种自组织，状态转换的各层次并非互相分离，而是实时透明传递及处理信息的。

搞人工智能或类脑智能有一个目的性问题，即我们不是构造完全相同的人脑，而是仿脑。两者的本质区别在于人脑有历史积淀，已经形成相对固定模式，而仿脑可以跳出这种相对固定模式，完全一样反而不能在某一方面超越，仅仅是一样而已，类脑只有超越才能体现更大价值。

模拟人脑智能，只要与人脑具有逻辑同构性，就可以在记忆、逻辑推理、知识的条理化方面超越人脑，因为人脑要受环境及教育程度的限制，而仿脑没有这样的限制，它可以集中全人类的智慧于一体！因而超过单一个体人脑是理所当然的事！仿脑最重要的一点就是具有超强的记忆能力，基于逻辑内涵与外延的逻辑一致性，仿脑可以把人类历史上、及所有人的知识集中在一起，且比人脑的信息处理速度更快、更精确、更具系统性！因而比人处理信息的能力更强！

99-702, 神经形态计算中的动态可靠性管理

Shihao Song、Jui Hanamshet、Adarsha Balaji、Anup Das、Jeffrey L. Krichmar、Nikil D. Dutt、Nagarajan Kandasamy、Francky Catthoor 2021.5.5

加州大学、德雷塞尔大学、比利时微电子研究中心

神经形态计算系统使用非易失性存储器 (NVM) 来实现高密度和低能耗的突触存储。操作 NVM 所需的升高的电压和电流会导致每个神经元和硬件中的突触电路的 CMOS 晶体管老化, 从而使晶体管的参数偏离其标称值。激进的设备扩展会增加功率密度和温度, 从而加速老化, 对神经形态系统的可靠运行造一成挑战。现在的以可靠性为导向的技术 (假设最坏的工作条件) 以固定的时间间隔周期性解除硬件中所有神经元和突触电路的工作负载, 而实际上并未在运行时跟踪它们的老化情况。为了减轻这些电路的压力, 必须中断正常操作, 这会在脉冲产生和传播中引入等待时间, 从而影响脉冲间隔和性能, 例如精度。通过设计智能运行时管理器 (NCRTM), 我们提出了一种新的体系结构技术, 以缓解神经形态系统中与衰老相关的可靠性问题, 该管理器可动态响应神经元和突触电路在 CMOS 晶体管中的短期衰老, 从而缓解神经元和突触电路的压力。该管理器在执行机器学习工作负载期间, 以满足可靠性为目标。NCRTM 仅在绝对必要时才对这些电路施加压力, 否则会通过在关键路径之外安排压力操作来降低性能影响。我们使用神经形态硬件上的最新机器学习工作负载来评估 NCRTM。结果表明, NCRTM 显著提高了神经形态硬件的可靠性, 对性能的影响很小。

100-703, SpikE: 基于脉冲的多关系图数据嵌入

Dominik Dold、Josep Soler Garrido 2021.4.27

西门子

尽管最近成功地将基于脉冲的编码与错误反向传播算法相协调，但脉冲神经网络仍主要应用于感官处理产生的任务，这些任务在视觉或听觉数据等传统数据结构上运行。可以在工业和研究领域得到广泛应用的丰富数据表示形式就是所谓知识图谱-一种基于图的结构，其中实体被描述为节点，而它们之间的关系被描述为边缘。可以使用知识图谱嵌入算法在这些信息密集的环境中进行上下文感知的预测。我们提出了一种基于脉冲的算法，其中图中的节点由神经元种群的单个脉冲时间和种群之间的峰值时间差的关系表示。学习这样的基于脉冲的嵌入仅需要有关脉冲时间和脉冲时间差的知识，就可以与最近提出的用于训练脉冲神经网络的框架兼容。提出的模型可以轻松地映射到当前的神经形态硬件系统，从而将对知识图谱的推理移入这些体系结构的领域，从而为该技术打开了一个有前途的工业应用领域。

101-707, 2021 年神经形态计算与工程路线图

丹麦科技大学，亚琛科技大学，2021.5.12

如今，基于冯-诺依曼架构的现代计算已成为一门成熟的前沿科学。在这种体系结构中，处理和存储单元被分为单独的模块，可以密集且连续地交换数据。这种数据传输造成了很大一部分功耗。下一代计算机技术有望解决这种超大规模的计算问题，即使这些未来的计算机具有令人难以置信的强大功能，但是如果它们基于冯-诺依曼类型的体系结构，它们将消耗 20 到 30 兆瓦的功率，并且不会像我

们的大脑那样具有内在的物理内置能力来学习或处理复杂的非结构化数据。神经形态计算系统旨在满足这些需求。人脑使用 20W 和 1.2L 的体积每秒执行约 10~15 次计算。通过从 42 物理学中获得启发，新一代计算机的功耗可以比传统处理器低得多，可以利用集成的非易失性存储器和逻辑，并且可以明确地设计为在复杂和非结构化数据的情况下支持动态学习。在其潜在的未来应用中，商业、医疗保障、疾病和病毒的传播控制在社会层面可能是最有影响力的。该路线图设想了神经形态材料在前沿技术中的潜在应用，并将重点放在人工神经网络的设计和制造上。该路线图的内容将突出这一活动的跨学科性质，该活动的灵感来自生物学，它从生物学、物理学、数学、计算机科学和工程学中获得灵感。这将为在许多技术相关领域探索和整合当前和未来应用背后的新技术提供路线图。

102-711, 系统软件在神经拟态计算能耗管理中的作用

德雷克塞尔大学 2021. 3. 22

最近，很多计算社区引入了将诸如 DYNAP 和 Loihi 之类的神经拟态计算系统以提高机器学习程序的性能和能效，尤其是那些使用 Spiking Neural Network (SNN) 实现的程序。用于神经拟态系统的系统软件的作用是将大型机器学习模型（例如具有许多神经元和突触）聚类并将这些聚类映射到硬件的计算资源。本文考虑了神经元和突触消耗的功率，以及在互连上传递尖峰时所消耗的能量，从而制定了神经拟态硬件的能量消耗。文中首先评估了系统软件在管理神经拟态系统能量消耗中的作用。作者制定了一种简单的映射方法，将神经元和突触放置到计算资源上以减少能耗。最终文本用 10 个机器学习应用程序评估了提出的方法，并证明了所提出的映射方法可显著减少神经拟态计算系统的能耗。

103-712, 稀疏脉冲梯度下降

帝国理工学院 2021. 5. 18

由于神经形态计算设备的能耗低，人们越来越喜欢在其上模拟脉冲神经网络（SNN）。最近的进展使训练 SNN 的准确性达到了与传统人工神经网络（ANN）竞争的地步。同时在神经形态硬件上运行时具有能源效率。但是，训练 SNN 的过程仍然基于最初为 ANN 开发的密集张量操作，该操作不利用 SNN 的时空稀疏性质。我们在这里介绍了第一种稀疏 SNN 反向传播算法，该算法可实现与当前技术水平相同或更高的准确性，同时显著提高了速度并提高了存储效率。展示我们的方法在不同复杂度的真实数据集（MNIST，神经语言 MNIST 和 Spiking Heidelberg Digits）上的有效性，该方法在不减少精度的情况下实现了高达 70 倍的后向加速并提高 40% 的内存效率。

104-716, 具有超分辨率忆阻器交叉开关的模拟神经计算

IEEEA. P. James, L. O. Chia, 2021. 5. 10

忆阻器交叉开关阵列广泛用于内存中和神经形态计算应用，但是，忆阻器器件存在非理想状态，会让导电状态发生变化，因此随着器件的老化，将其编程为所需的模拟电导的一组稳定电导值。交叉开关中可用于节点的电导级别的数量被定义为交叉开关的分辨率。

本文提出了一种通过建立超分辨率忆阻器交叉开关来提高分辨率的技术，该超分辨率忆阻器交叉开关的节点具有多个忆阻器，以生成具有唯一电导值的 r -单纯形序列。电导值的范围和数量越宽，交叉开关的分辨率越高。这在构建模拟神经网络（ANN）层时特别有用，事实证明，这被证明是实现神经形态计算中形成神经网络

络层的一种可行方法。

105-722, 时态神经网络在线学习的微体系结构实现框架

Harideep Nair, John Paul Shen, James E. Smith, 卡内基梅隆大学, 2021. 5. 27

时态神经网络 (TNN) 是含有脉冲的神经网络, 类似于哺乳动物的新皮层, 它使用时间作为资源来表示和处理信息。与采用单独的训练和推理阶段的计算密集型深度神经网络相反, TNN 具有极高效率的在线增量 / 连续学习能力, 并且是构建边缘本机感觉处理单元的极佳候选者。这项工作提出了使用标准 CMOS 来实现 TNN 的微体系结构框架。提出了三个关键构建块的门级实现: ①多突触神经元, ②多神经元列, 以及③基于峰值定时依赖可塑性 (STDP) 的无监督和监督在线学习算法。TNN 微体系结构体现在一组特性缩放方程式中, 可用于评估任何 TNN 设计的门数、面积、延迟和功耗。论文提出了拟议设计的合成后果 (在 45nmCMOS 中), 并证明了其在线增量学习能力。

106-724, 使用尖峰神经网络对基于事件的光流进行自监督学习

代尔夫特理工大学 Federico Paredes-Valles 等

神经形态传感和计算有望实现高能效和高带宽传感器处理。神经形态计算的一个主要挑战是, 由于离散脉冲和复杂的神经元动力学, 传统人工神经网络 (ANN) 的学习方法不能直接转移到脉冲神经网络 (SNN)。因此, SNN 尚未成功应用于复杂的大规模任务。在本文中, 我们专注于从基于事件的相机输入进行光流估计的自监督学习问题, 并研究最先进的 ANN 训练管道所需的变化, 以便成功地转换为 SNN。我们首先修改输入事件表示, 以使用最少的显式时间信息对更小的时间片进行编

码。因此，使网络的神经元动力学和循环连接随着时间的推移能够整合信息。此外，我们重新制定了基于事件的光流的自监督损失函数以改善其特性。我们使用提议的管道对各种模型的循环神经网络和 SNN 进行实验。关于 SNN，我们研究了参数初始化和优化、替代梯度形状和自适应神经元机制等元素的影响。结果发现初始化和代理梯度宽度在实现稀疏输入学习方面起着至关重要的作用，而包含自适应性和可学习的神经元参数可以提高性能。实验表明，所提出的 ANN 和 SNN 的性能与以自我监督方式训练的当前最先进的 ANN 的性能相当。

107-725, 探索神经拟态学习规则的库

芝加哥大学 RickStevens 等, 2021.5.1

神经拟态计算领域正处于积极探索的时期。虽然已经开发出了许多模拟神经元动力学、将深度神经网络转换为 SNN 模型的工具，但用于神经拟态学习的规则的通用软件库仍未得到充分的探索与开发。为神经拟态网络设计新的学习规则库具有非常大的挑战性，其范围从编码方法到梯度近似，从模仿贝叶斯大脑的群体方法到部署在忆阻器交叉开关上的约束学习方法。为了填补这一方面的空白，本文提出了一个模块化、可扩展的库 NeKo，专注于帮助设计新的学习算法，并在三个示例案例中展示了 NeKo 的效用，包括在线本地学习、概率学习和模拟设备上的学习。实验结果表明，NeKo 可以复制最先进的算法，并且其在准确性和速度方面的表现明显优于其他算法。此外它还提供了包括梯度比较在内的工具，可以帮助开发新的算法变体。NeKo 是一个开源 Python 库，支持 PyTorch 和 TensorFlow 后端。

108-729, MERLOT: 多模态神经脚本知识模型

华盛顿大学 Allen 人工智能研究所, 2021. 6. 4

人类观察事件的视觉上下文, 进行跨时间的多模态推理, 对过去、现在和未来做出推断, 进而理解世界。

本文提出 MERLOT, 通过观看数以百万计的 YouTube 视频和转录语音, 来学习多模态脚本知识的模型——以一种完全无标记、自监督的方式。通过对帧级 (空间) 和视频级 (时间) 目标的混合预训练, 模型不仅学会了将图像与时间上对应的词语相匹配, 还学会将全局发生的事件与时间相联系, 表现出强大的开箱即用的时间常识表示, 微调后在 12 个不同的视频问答保证数据集上实现了最先进的性能, 能很好地迁移到静态图像世界, 允许模型推理视觉场景背后的动态上下文。在视觉共感推理中, MERLOT 以 80.6% 的准确率回答问题, 比类似规模的先进模型高出 3% 以上, 即使是那些大量使用辅助监督数据 (如物体边框) 的模型。消融分析显示了以下的互补重要性: 在视频与静态图像上进行训练; 扩大预训练视频语料库的规模和多样性; 使用不同的目标, 鼓励从识别到认知层面的全栈多模态推理。

109-738, Spiking-GAN: 使用 Time-To-First-Spike 编码的 Spiking 生成对抗网络

Vineet Kotariya 等, 印度理工大学

脉冲神经网络 (SNN) 在以节能的方式解决深度学习问题方面显示出了巨大的潜力。但是, 它们仍然仅限于简单的分类任务。在本文中, 我们提出了 Spiking-GAN 这是第一个基于脉冲的生成对抗网络 (GAN)。它采用一种称为首次脉冲时间编码的时间编码方案。我们使用时域中的近似反向传播来训练它。我们的网络使用非

常高不应期的简单积分和激发 (IF) 神经元, 以确保每个神经元最多有一个脉冲。这使得模型比基于脉冲率的系统稀疏得多。与之前的工作相比, 修改后的时间损失函数称为 “Aggressive TTFS”, 将网络推理时间提高了 33% 以上, 并将网络中的脉冲数量减少了 11% 以上, 实验表明, 使用这种方法在 MNIST 数据集上训练网络时, 我们可以生成高质量的样本, 从而证明了该框架在解决脉冲域中此类问题的潜力。

110-742, 深度神经网络中的时间稀疏性训练在视频处理中的应用

比利时微电子研究中心, 2021. 7. 15

激活稀疏性提高了稀疏感知神经网络加速器的计算效率和资源利用率。由于 DNN 中的主要操作是具有权重的激活的乘法累加 (MAC) 以计算内积, 因此跳过 (至少) 两个操作数之一为零的操作可以使推理在延迟和功率方面更有效。激活的空间稀疏化是 DNN 文献中的一个热门话题, 并且已经建立了几种方法来使 DNN 偏向于它。另一方面, 时间稀疏性是受生物启发的脉冲神经网络 (SNN) 的固有特征, 神经形态处理利用它来提高硬件效率。引入和利用时空稀疏性是 DNN 文献中很少探讨的主题, 但与 DNN 的趋势完美共鸣, 从静态信号处理转向更多流信号处理。为了实现这一目标, 在本文中, 我们引入了一个新的 DNN 层 (称为 Delta 激活层), 其唯一目的是在训练期间促进激活的时间稀疏性。Delta 激活层将时间稀疏性转换为空间激活稀疏性, 以便在硬件中执行稀疏张量乘法时加以利用。通过在训练期间采用增量推理和 “通常的” 空间稀疏化启发式方法, 所得模型不仅学习利用空间而且利用时间激活稀疏性 (对于给定的输入数据分布)。人们可以在原版训练或细化阶段使用 Delta 激活层。我们已经实现了 Delta 激活层作

为标准 Tensorflow-Keras 库的扩展，并将其应用于在人类动作识别 (UCF101) 数据集上训练深度神经网络。我们发现，激活稀疏性提高了近 3 倍，在长时间训练后模型准确性的损失是可恢复的。

111-745, 用于提高神经网络对数据质量问题的鲁棒性的调制层

华盛顿大学, 2021. 7. 19

数据质量是机器学习中的一个常见问题，特别是在医疗保健等高风险环境中。缺失数据会影响复杂模式中的精度、校准和特征属性。开发人员经常在精心策划的数据集上训练模型，以最小化丢失的数据偏差；但是，这降低了此类模型在生产环境（如实时医疗记录）中的可用性。因此，使机器学习模型对缺失数据具有鲁棒性是实际应用的关键。一些分类器自然地处理缺失数据，而另一些分类器，如深度神经网络，则不是针对未知值设计的。我们提出了一种新的神经网络修正方法来减轻缺失数据的影响。这种方法的灵感来自于由生物神经网络进行的神经调节。我们的建议将完全连接层的固定权重替换为每个输入的附加输入（可靠性得分）函数，模仿大脑皮层基于其他数据的上下权重输入的能力。利用多层感知器与主任务联合学习调制函数。我们在多重分类、回归和插补问题上测试了我们的调制全连接层，它要么提高了性能，要么产生了与传统神经网络结构相类似的性能，将可靠性连接到输入。具有调制层的模型通过在评估时引入额外的缺失，对数据质量的降低更具鲁棒性。这些结果表明，通过调制全连接层明确说明信息质量的降低可以使人工智能系统在实时环境中部署。

112-752, 基于数据流将脉冲神经网络合成到神经形态硬件

ShihaoSong 等, 德雷塞尔大学, 2021.8.4

脉冲神经网络 (SNN) 是一种新兴的计算模型, 它使用事件驱动的激活和仿生学习算法。基于 SNN 的机器学习程序通常在基于 tile 的神经形态硬件平台上执行, 其中每个 tile 由一个称为 crossbar 的计算单元组成, 它映射程序的神经元和突触。然而, 在现成的神经形态硬件上合成此类程序具有挑战性。这是因为硬件的固有资源和延迟限制会影响模型性能 (例如准确性) 和硬件性能 (例如吞吐量)。我们提出了 DFSynthesizer, 一种将基于 SNN 的机器学习程序合成到神经形态硬件的端到端框架。提议的框架分四个步骤工作。首先, 它分析机器学习程序并使用代表性数据生成 SNN 工作负载。其次, 它划分 SNN 工作负载并生成适合目标神经形态硬件纵横的集群。第三, 它利用同步数据流图 (SDFG) 的丰富语义来表示集群 SNN 程序, 允许在关键硬件约束方面进行性能分析, 例如交叉开关的数量、每个交叉开关的尺寸、瓦片上的缓冲区空间和瓦片通信带宽。最后, 它使用一种新颖的调度算法在硬件的交叉开关上执行集群, 保证硬件性能。我们使用 10 个常用的机器学习程序评估 DFSynthesizer。我们的结果表明, 与当前的映射方法相比, DFSynthesizer 提供了更优秀的性能保证。

113-753, 节能随机游走计算的神经拟态缩放优势

J. Darby Smith, 神经探索与研究实验室, 桑迪亚国家实验室, 2021.7.27

受人脑运作方式启发的神经拟态计算 (NMC) 方法可以从根本上改进现有的计算方式。大多数旨在在人造硬件中复制大脑计算结构和架构的 NMC 研究都集中在人工智能上, 很少有人探索这种受大脑启发的硬件是否可以提供超越认知任务的

价值。本文证明了尖峰神经拟态架构的高度并行性和可配置性使它们非常适合通过离散时间马尔可夫链实现随机游走算法。这种随机游走在蒙特卡罗方法中很有用，蒙特卡罗方法代表了解决各种数值计算任务的基本计算工具。此外，本文还展示了涉及一类随机微分方程的概率解决方案的数学基础如何利用这些模拟为一系列广泛适用的计算任务提供解决方案。同时，本文还发现 NMC 平台在足够的规模下可以大大降低高性能计算 (HPC) 平台的能源需求。

114-758, 具有相变记忆突触的 SRNN 的在线训练

香港大学, 2021. 8. 4

SRNN 因其丰富的时间动态和稀疏处理而成为解决各种复杂认知和运动任务的有效工具。然而，由于即使在权重分辨率有限的情况下，仍然缺乏本地的、硬件友好的学习机制来解决时间信用分配问题并确保稳定的网络动态，在专用的神经拟态硬件上训练 SRNN 仍然存在着诸多争议。如果使用忆阻器件进行内存计算来解决冯诺依曼瓶颈问题，将会显著增加 SRNN 的计算和工作内存的可变性。为了应对这些挑战并在忆阻 SRNN 中实现在线学习，本文提出了一个基于准确和全面的相变存储器 (PCM) 设备模型的差分架构交叉阵列的仿真框架，使用最近提出的电子道具学习规则训练了一个 SRNN，并利用模拟框架模拟其权重。尽管 e-prop 局部接近理想的突触更新，但由于大量的 PCM 非理想性，很难在忆阻基板上实现更新。同时，本文还比较了几种广泛适用的权重更新方案，这些方案主要旨在应对这些设备的非理想情况，并证明累积梯度可以实现在忆阻基板上在线和有效训练 SRNN。

115-762, 神经网络模型的记忆容量

哥伦比亚大学, 2021. 8. 17

记忆是一种复杂的现象, 涉及到几种不同的机制。这些机制在不同的空间和时间层面上运作。本章重点介绍了理论框架和数学模型, 这些模型是为了理解这些机制是如何被编排来存储、保存和检索大量记忆的。特别是, 本章回顾了关于记忆容量的理论研究, 研究人员在研究中估计了可存储记忆的数量如何与神经回路中神经元和突触的数量成比例。记忆容量取决于突触的复杂性、表征的稀疏性、记忆之间的时空相关性以及记忆提取的具体方式。当突触只能以有限的精度进行修改时(如生物突触), 复杂性很重要, 而稀疏性可以极大地提高记忆容量, 在记忆结构(相互关联)时尤其有益。本章讨论的理论工具可以用来确定记忆存储、保存和检索的重要计算原理, 并为设计和解释记忆实验提供指导。

116-776, 用于序列学习的具有改进的固有循环动力学的尖峰神经网络

普渡大学, 2021. 9. 4

具有 LIF 神经元的 SNN 以事件驱动的方式运行, 并存储内部状态以随时间保留信息, 为在边缘设备上的节能神经拟态计算提供机会。然而, 许多关于 SNN 的代表性工作并没有完全证明其固有的循环(保留过去信息的膜电位)对于顺序学习的有效性。大多数作品通过速率编码及时人为扩展输入表示来训练 SNN 识别静态图像。本文展示了 SNN 可以针对顺序任务进行训练, 并提出对 LIF 神经网络的修改, 使内部状态能够学习长序列并使其固有的递归能够适应梯度消失问题。然后, 本文呢作者还开发了一个训练方案来训练所提出的具有改进的固有递归动力学的 SNN, 该方案允许尖峰神经元产生多位输出(与二进制尖峰相反), 这有

助于减轻尖峰神经元激活函数的导数与用于克服尖峰神经元不可微性的替代导数之间的不匹配。实验结果表明，在 TIMIT 和 Libri Speech 100h 数据集上提出的 SNN 架构产生的精度与 LSTM 相当（分别在 1.10% 和 0.36% 以内），但参数比 LSTM 少 2 倍。与 GRU 相比（GRU 通常被认为是 LSTM 的轻量级替代品），稀疏 SNN 输出还分别在 TIMIT 和 Libri Speech 100h 数据集上节省了 10.13 倍和 11.14 倍的乘法运算。

117-781, 生物启发的语音情感识别

Reza Lotfidereshgi 等, Sherbrooke 大学, 2021.11.15

传统的基于特征的分类方法不能很好地应用于语音情感的自动识别，主要是因为识别说话人情感状态所需的精确的光谱和韵律特征集尚未确定。

本文提出了一种直接对语音信号进行处理的方法，从而避免了特征提取的困难步骤。此外，该方法结合了人类语音产生的经典源滤波器模型和最近引入的液态机（LSM）的优点，后者是一种受生物启发的尖峰神经网络（SNN）。首先分离语音信号的源和声道分量，并将其转换为感知相关的频谱表示。然后由两个神经元库分别处理这些表示。该方法在柏林情感语音数据库（Emo-DB）上具有很好的分类性能。

118-786, 一种通过 FPGA 上新兴的神经编码加速脉冲神经网络的端到端框架

Daniel Gerlinghoff 等, 香港中文大学（深圳）, 2021.11.19

内容：编译器框架对于广泛使用基于 FPGA 的深度学习加速器至关重要。它们允许不熟悉硬件工程的研究人员和开发人员利用特定领域逻辑获得的性能。传统人

工神经网络存在多种框架。然而，在创建针对脉冲神经网络（SNN）优化的框架方面并未投入太多研究工作。这种新一代神经网络对于在具有严格功率和资源限制的边缘设备上部署 AI 变得越来越有趣。我们的端到端框架 E3NE 可自动为 FPGA 生成高效的 SNN 推理逻辑。它基于 PyTorch 模型和用户参数，应用各种优化并评估基于脉冲的加速器的固有优缺点。多级并行和新兴神经编码方案的使用导致效率优于以前的 SNN 硬件实现。对于类似的模型，E3NE 使用不到 50% 的硬件资源和 20% 的功耗，同时将延迟降低了一个数量级。此外，可扩展性和通用性允许部署大规模 SNN 模型 AlexNet 和 VGG。

119-787, 使用 Loihi-2 进行高效的神经形态信号处理

Garrick Orchard 等, 英特尔, 2021. 11. 5

内容: 神经形态计算中使用的受生物启发的尖峰神经元是带有动态状态变量的非线性滤波器——与深度学习中使用的无状态神经元模型非常不同。英特尔神经形态研究处理器 Loihi 2 的下一个版本支持多种具有完全可编程动态的状态脉冲神经元模型。在这里, 我们展示了先进的尖峰神经元模型, 可用于在仿真 Loihi 2 硬件上的模拟实验中有效地处理流数据。在一个例子中, 共振和激发 (RF) 神经元用于计算短时傅里叶变换 (STFT), 其计算复杂度与传统 STFT 相似, 但输出带宽比传统 STFT 少 47 倍。在另一个例子中, 我们描述了一种使用时空 RF 神经元的光流估计算法, 与传统的基于 DNN 的解决方案相比, 该算法需要的运算量少 90 倍以上。我们还展示了利用反向传播训练用于音频分类任务的 RF 神经元的有希望的初步结果。

最后, 我们证明了 Hopf 谐振器 (RF 神经元的一种变体) 的级联复制了耳蜗的新

特性，并激发了一种有效的基于尖峰的频谱图编码器。

120-797, 面向节能嵌入式神经拟态计算多核 BigLittleuBrain 设计

M.L. Varshika, 比利时德雷塞尔大学, 2021. 11. 23

内容: 随着嵌入式系统中基于脉冲的深度学习推理应用的增加, 这些系统倾向于集成神经拟态加速器, 例如 μ 提高能源效率的大脑。我们提出一个 μ 基于大脑的可扩展多核神经拟态硬件设计, 可加速脉冲深度卷积神经网络 (SDCNN) 的计算。为了提高能源效率, 内核在其神经元和突触容量方面被设计为异构 (大内核比小内核具有更高的容量), 并且它们使用并行分段总线互连进行互连, 与相比, 这导致更低的延迟和能量到传统的基于网格的片上网络 (NoC)。我们提出了一个名为 SentryOS 的系统软件框架, 将 SDCNN 推理应用程序映射到所提出的设计。SentryOS 由一个编译器和一个运行时管理器组成。编译器利用 big 和 little 的内部架构将 SDCNN 应用程序编译成子网络 μ 脑核。运行时管理器将这些子网络调度到核心上并流水线化它们的执行以提高吞吐量。我们使用五个常用的 SDCNN 推理应用程序评估了所提出的大、小众核神经拟态设计和系统软件框架, 并表明所提出的解决方案降低了能量 (37%到 98% 之间), 减少了延迟 (9%到 25%之间), 并增加应用程序吞吐量 (在 20%到 36%之间)。我们还表明, SentryOS 可以很容易地扩展到其他脉冲神经拟态加速器。

121-810, 基于 STDP 的尖峰神经网络监督学习算法

Zhanhao Hu 等, 2022. 3. 7

清华大学,

内容: 与基于速率的人工神经网络相比, SNN 提供了更具生物学合理性的模型为了大脑。但他们如何进行监督学习仍然是个谜。作者提出了一种基于尖峰时间依赖可塑性 (STDP) 的有监督学习算法, 用于分层 SNN 由漏泄整合和激发 (LIF) 神经元组成。A. 时间窗是为突触前神经元设计的, 只有棘波在此窗口中, 参与 STDP 更新过程。模型是在 MNIST 数据集上训练。分类精度的方法是具有类似结构的多层感知器 (MLP) 的标准的反向传播算法。作者描述了一种基于 STDP 的 SNN 监督学习算法, 并得到了相应的结果在 MNIST 分类任务中取得了良好的结果。准确度接近具有类似体系结构的 MLP, 这表明了该方法的有效性算法。与现有的 SNN 训练算法相比, 本文提出了一种新的 SNN 训练算法算法已经取得了相互竞争的结果。该算法表明, 生物神经元可能不会改变它们的突触一直在 STDP 规则下。STDP 仅在监管机构批准时生效信号被应用。此外, 该算法表明, 并不是所有的峰值突触前神经元参与突触的 STDP 学习过程。可能存在一个时间窗口, 只有道琼斯指数在这场胜利中的峰值才会被计算在内。但需要生化证据来验证这些预言。

122-811, 类神经拟态计算的替代梯度

BenjaminCramer 等, 2022. 1. 14

德国海海德堡大学, 瑞士弗雷德里克. 米歇尔生物医学研究所

内容: 为了以低代谢成本快速处理时间信息, 生物神经元以模拟和的形式整合为输入, 但及时与脉冲、二进制事件通信。类神经拟态硬件使用相同的原理来模拟

具有出色能效的脉冲神经网络。然而，由于设备不匹配和缺乏有效的训练算法，在此类硬件上实例化高性能脉冲网络仍然是一项重大挑战。替代梯度学习已成为一种有前景的脉冲网络训练策略，但尚未证明其对类神经拟态系统的适用性。在这里，我们使用循环方法演示了 BrainScaleS-2 类神经拟态系统的替代梯度学习。我们证明，学习可以自我纠正设备不匹配，导致在视觉和语音基准上具有竞争力的脉冲网络性能。我们的网络显示平均每个隐藏神经元和输入不到一个脉冲的稀疏脉冲活动，以高达每秒 85,000 帧的速率执行推理，消耗不到 200 mW。总之，我们的工作为类神经拟态硬件上的低能量脉冲网络处理设定了几个基准，并为未来的片上学习算法铺平了道路。

123-822, 自适应尖峰神经网络的精确平均场模型

LiangChen 等, 2022. 3. 16,

滑铁卢大学

内容：具有适应性的棘突神经网络已被证明能够重现广泛的神经活动，包括突发群体爆发和棘突同步，这是大脑紊乱和正常功能的基础。从尖峰神经网络导出的精确平均场模型非常有价值，因为这样的模型可以用来确定单个神经元和网络参数如何相互作用，从而产生宏观网络行为。在本文中，作者推导并分析了具有尖峰频率自适应的神经网络的一组精确平均场方程。具体来说，作者的模型是一个伊兹克维奇神经网络，其中每个神经元由一个二维系统建模，该系统由一个二次积分和火灾方程加上一个实现尖峰频率自适应的方程组成。之前的工作是为这种类型的网络推导平均场模型，依赖于适应变量足够慢的动力学假设。然而，这种近似并没有成功地在宏观描述和现实神经网络之间建立精确的对应关系，尤

其是在适应时间常数不大的情况下。挑战在于如何通过包含自适应变量的平均场表达式来实现一组封闭的平均场方程。作者通过使用洛伦兹-安萨兹结合力矩闭合方法来解决这一挑战，从而得到热力学极限下的平均场系统。由此产生的宏观描述能够定性和定量地描述神经网络的集体动力学，包括紧张性放电和爆发之间的转换。

124-823, 节能随机游走计算的神经拟态缩放优势

J. Darby Smith 等, 2022. 2. 14,

美国桑迪国家实验室, 神经探索与研究实验室

内容: 神经拟态计算的目的是在合成硬件中复制大脑的计算结构和架构, 它通常专注于人工智能应用。人们较少探讨的是, 这种受大脑启发的硬件能否提供超越认知任务的价值。在这里, 我们展示了脉冲神经拟态结构的高度并行性和可配置性, 使其非常适合通过离散时间马尔可夫链实现随机游动。这些随机游动在蒙特卡罗方法中很有用, 蒙特卡罗方法是解决广泛数值计算任务的基本计算工具。通过使用 IBM 的 TrueNorth 和 Intel 的 Loihi 神经形态计算平台, 我们表明, 与传统方法相比, 用于生成扩散的随机游走近似的神经拟态计算算法在节能计算方面具有优势。我们还表明, 我们的神经拟态计算算法可以扩展到更复杂的跳跃-扩散过程, 这些过程在金融经济学、粒子物理和机器学习等一系列应用中都很有用。



敬请关注联盟微信公众号
COPU开源联盟



扫描二维码
获取往期资料

中国开源软件推进联盟秘书处

电话：+86 010-88558999

联盟公共邮箱：office@copu.org.cn

联盟官网：<http://www.copu.org.cn>

地址：北京市海淀区紫竹院路66号赛迪大厦18层