智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

认知机器的结构和激活

李德毅, 张天雷, 韩威, 海丹, 鲍泓, 高洪波

引用本文:
李德毅, 张天雷, 韩威, 等. 认知机器的结构和激活[J]. 智能系统学报, 2024, 19(6): 1604-1613.
LI Deyi, ZHANG Tianlei, HAN Wei, et al. Structure and activation of cognitive machines[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(6): 1604-1613.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202409024

DOI: 10.11992/tis.202409024

认知机器的结构和激活

李德毅[1], 张天雷[2], 韩威[3], 海丹[3], 鲍泓[4], 高洪波[5]

（1. 清华大学 信息科学技术学院, 北京 100084; 2. 北京主线科技有限公司, 北京 100083; 3. 北京中科原动力科技有限公司, 北京 100085; 4. 北京联合大学 机器人学院, 北京 100101; 5. 中国科学技术大学 信息科学技术学院, 安徽 合肥 230026）

摘　要: 从物理学的角度理解人类认知已经成为当今人工智能面临的核心挑战。本文分析了计算机通用结构中, **孤立计算、忽视记忆**和**孤立思维、忽视具身**的局限性。以驾驶认知为例, 提出了包括感知、思维、行为在内的认知机器的通用结构组成。它区别于计算机的架构, 也区别于杨立昆的“世界模型”和李飞飞的“空间模型”, **增加了记忆组块, 用人工智痕元胞作为神经元细胞的镜像, 用智痕元胞网络构成思维软构体, 实现记忆智能的生成、调控和提取。**物质硬构体可采用 CPU、DPU、GPU、TPU、FPGA、SSD、搜索引擎等并行处理单元, 其系统是分布的、并行的、异构的。一旦加电获得能量, 机器就不再是死物质, 认知核中的思维软构体和物质硬构体经过一番纠缠, 机器被激活。激活后的机器赖负熵为生, 进入和物理世界具身交互的认知状态。**机器认知像人又不像人, 宕机后可再激活, 能自主感知、思维、决策和行为, 可交互, 会学习, 自纠错, 自成长。**该结构既可用于构造数字虚拟机器人, 也可用于构造替代人类劳动岗位的、千姿百态具身的机器人, 使得人类能够迅速进入人机共生的智能时代。

关键词: 认知核; 智痕元胞网络; 纠缠; 记忆智能; 具身智能

中图分类号: TP18　　**文献标志码:** A　　**文章编号:** 1673−4785(2024)06−1604−10

中文引用格式: 李德毅, 张天雷, 韩威, 等. 认知机器的结构和激活 [J]. 智能系统学报, 2024, 19(6): 1604–1613.

英文引用格式: LI Deyi, ZHANG Tianlei, HAN Wei, et al. Structure and activation of cognitive machines[J]. CAAI transactions on intelligent systems, 2024, 19(6): 1604–1613.

Structure and activation of cognitive machines

LI Deyi[1], ZHANG Tianlei[2], HAN Wei[3], HAI Dan[3], BAO Hong[4], GAO Hongbo[5]

(1. School of Information Science and Technology, Tsinghua University, Beijing 100084, China; 2. Trunk Technology Corporation Ltd., Beijing 100083, China; 3. AIforce Technology Corporation Ltd., Beijing 100085, China; 4. School of Robotics, Beijing Union University, Beijing 100101, China; 5. School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: Understanding human cognition from a physical perspective is a core challenge for artificial intelligence (AI). This study analyzes the limitations of isolated computing, neglecting memory and isolated thinking, and overlooking embodied intelligence within the general structure of computers. In the driving cognition scenario, we propose a general structure of cognitive machines encompassing perception, thinking, and behavior, which differs from the computer architecture, as well as from Yann LeCun's "world model," and Li Feifei's "space model." It introduces memory blocks and employs artificial intelligent trace meta-cells as mirrors of neural cells, and intelligent trace meta-cells networks are used to form the thinking soft-structured ware to realize the generation, regulation, and extraction of memory intelligence. The hard-structured ware can use parallel processing units such as CPU, DPU, GPU, TPU, FPGA, and search engines. Therefore, the system must be heterogeneous. Once powered on, the machine is no longer a pile of inert matter; it will be activated through the entanglement between the thinking soft-structured ware and the material hard-structured ware in the cognitive nucleus. The activated machine thrives on negative entropy and enters a cognitive state of interaction with the physical world. Machine cognition is both human-and non-human-like; it can be reactivated after downtime, achieving independent perception, cognition, decision-making and behavior, interaction, learning, and self-growth. This structure can be used to construct digital virtual robots and intelligent agents with a myriad of shapes to replace human labor, rapidly propelling humanity into an intelligent era of man–machine symbiosis.

Keywords: cognitive nucleus; intelligent trace meta-cell networks; entanglement; memory intelligence; embodied intelligence

1　计算机体系结构的局限性

1.1　从计算机器的结构谈起

随着类脑计算的蓬勃发展, 从物理学的角度理解人类认知已经成为当今人工智能面临的核心难题。当今的计算机本质上实现的是机械的、电子的、非生命的计算装置。它是能实证的、可量化的, 可以用逻辑学的方法证明或者计算, 如数值计算、优化计算、符号逻辑、谓词演算、定理证

收稿日期: 2024−09−18.

通信作者: 李德毅, E-mail: lidy@cae.cn.

明、概率计算等。早期的冯·诺依曼计算机由 CPU（控制单元和运算单元）、内存、外存和输入、输出组成。计算机体系结构更强调构成计算机系统的各个组件的内部结构及其相互关系，以及计算机系统软硬件之间的接口关系。它包括**指令集体系结构**和**微体系结构**两个层面，**指令集体系结构**是思维软构体和物质硬构体之间的界面，定义了处理器可以执行的指令集合（复杂指令集或者精简指令集）、数据类型、寄存器、内存访问方式、输入输出机制等。**微体系结构**是处理器内部的物理实现，即物质硬构体，它得益于固体物理学的研究成果，尤其是半导体芯片和集成电路技术，包括 CPU 内部的寄存器、数据路径、控制单元、缓存等组件。计算机体系结构还涉及支持多核处理、众核处理、包括 GPU 在内的异构处理单元等。**这样的物理装置究竟怎么完成人的计算和思维的呢?**

图灵和冯·诺依曼都是数学家和物理学家，他们发明的计算机器的结构，可以实现人的计算智能，甚至能够思维，但他们并不是生命科学家。同样，获得诺贝尔奖的著名物理学家薛定谔也不是生命科学家，但他的著作《生命是什么——活细胞的物理学观》，对生物学领域产生了重要影响。生命科学中发现 DNA 双螺旋结构的科学家，仍然不是生物学家，而是物理学家克拉克、沃森等，他们因此获得诺贝尔生理和医学奖。这就说明物理学对生物学的基础性作用，并诞生出一个十分有价值的交叉研究学科——生物物理学。

本文试图以人类认知为突破口，用“赖负熵为生”的生命观，来解释机器是如何被激活的，以及机器是如何思维和认知的，即认知物理学。用认知来弥合生物学和物理学之间的鸿沟，填补生物认知和机器认知之间、人的智能和人工智能之间“缺失的连接”。

1.2　计算机架构中缺失记忆的形成、调控和提取

机器认知和人的认知一样，存在 4 种基本模式：记忆驱动的经验模式、知识驱动的推理模式、联想驱动的创造模式以及假说驱动的发现模式。认知依赖记忆，记忆是难以计算的智能，它先于计算、约束计算，无需解释。**记忆在这 4 种基本模式中都发挥着不可或缺的作用。**当前情境下发生的动态的、不确定性的记忆提取，常常体现了选择性注意。**但是，受图灵“智能的本质就是计算”的局限，传统人工智能只能是计算机智能，体系结构中只有简单的存储，缺失记忆的生成、调控和提取的组织结构。**冯·诺依曼架构的计算机，核心是算力和算法，通过程序实现算法，利用算力完成运算，**它不可能执行任何未预先编程的活动。**而机器认知主要是依靠记忆，计算机中的存储远远不能覆盖记忆的丰富内涵，认知机器需要模拟人脑数百亿神经元和数百万亿突触组成的记忆网络才行。互联网协议的伟大之处在于将应用程序和内容服务环境与底层传输结构的特征分开，互联网搜索技术历经的 30 多年发展演化和 ChatGPT 大模型的成功，都证明了一个事实：可以把互联网看成是一个超级记忆网络，无论是根据语法、语义、语境或者语用进行搜索，**云计算或者生成式人工智能，是一个类似超级人脑的、动态的、不确定性的记忆网络修剪和提取过程，不同粒度的记忆就是不同尺度抽象了的网络拓扑和表达，是复杂网络的数据挖掘而已。所以，一定要把记忆的形成、调控和提取机制引入到认知机器的架构中去。**

1.3　计算机架构中缺失具身交互认知

曾经的计算机是一种开环设计，它根据特定的输入，通过程序运行完成计算，给出输出结果。**今年计算、明年计算，在这里计算、在那里计算，结果都一样，不具有空间定位在内的感知能力，不具有时空智能，也没有具身行为动作的存在，只有启动状态和目标状态。**要达到目标状态，其解决方案就是一个行动序列，确保机器能够从启动状态最终达到目标状态，只要知道了问题答案就认为是解决了问题。如果在解决计算问题的过程中用户需要干预，则可以通过预设的人机交互界面，用鼠标、键盘、甚至语音等手段“填入”预设规格的相关内容。当然，这类交互技术进步很快，越来越趋于自然。**因此，在计算机科学和技术领域，输入输出司空见惯，人机交互耳熟能详，但把持久地和外界环境交互作为一种认知手段，作为智能体的具身智能，却不多见。**然而，实体机器一旦具身有了感知、认知和行为能力，能够学习、创作、成长的时候，越来越多的个性化虚拟数字人、千姿百态的实体机器人就可以作为我们的智能代理，替代我们的工作岗位。这时候，思考人和机器关系中的基本问题——具身交互认知，就被提上议事日程了。**机器在物理空间表现出的具身交互智能，完全应该也完全可以和认知空间的计算智能媲美，成为新一代认知机器体系结构组成的重要部分，一定不能再缺失了。**

2　认知机器的体系架构

在分析了计算机体系结构的上述两大局限性之后，我们从介观切入来研究认知机器的体系架构。介观指介于宏观与微观之间的一种体系。我

们不从化学的原子水平或者物理学的分子水平的微观角度，也不从脑组织功能分区的宏观角度，而是在神经元细胞与神经元网络水平的尺度上研究脑认知及其模拟，在介观上**认知机器由感知、记忆、交互和计算等部件组成，**其中的感知、认知与行为都是双向互动的。交互系统里常常有定时定位定姿、语音文字、图形图像等多种传感器部件，它们承担与外界环境的交互，而具身行为交互常常是一个反馈自调节的过程。**记忆是智能之母，是认知机器的核心。记忆系统里有瞬时记忆、工作记忆和长期记忆 3 类组块。**区别于冯·诺依曼架构，也区别于杨立昆的“世界模型”和李飞飞的“空间模型”，**人脑神经元细胞是记忆的基本单元，是物质硬构体，**它会发生持久的物理、化学变化，在细胞水平上留下记忆的印迹、痕迹、残迹，智痕可深可浅，连接有强有弱，体现神经元细胞和突触的可塑性。

我们提出用“智痕元胞”作为寄生在神经元细胞之上的思维软构体，它是物质硬构体——脑神经元细胞的镜像或指代。记忆智能是智痕元胞网络的一种整体具象，不是哪一个智痕元胞单独决定的。这个整体具象会随着知识的增长不断地演化，类似人的认知在生命过程中的二次扩张和持久重塑能力。它存在有用于检索记忆的多索引并行机制，从与时俱进的、不同抽象尺度的记忆网络中挖掘出网络节点的抱团特性和层次结构，挖掘出相互关联的诸多概念、认知地图、概念树、知识点和知识图，还有这些知识图随着时间变化的知识谱。认知机器的架构充分利用自学习，形成、修饰、并巩固记忆。当然，记忆智能也需要计算智能来帮忙，提高认知的可解释性。

2.1　典型案例：驾驶认知和驾驶脑

机器认知的范畴多种多样，具身千姿百态。其中，驾驶认知在全社会最备受关注。以交通为例，无论是飞机、船舶还是车辆，**已经并正在完成着无人驾驶的三部曲。第一步：人主驾，机器辅助驾驶。**20 世纪 50 年代的先进驾驶辅助系统（advanced driving assistance system, ADAS），被认为是现代巡航控制系统的前身。**第二步：机器自动驾驶，人随时干预。**美国汽车工程师协会（SAE）把自动驾驶分成 5 个等级，利用自动控制技术，实现对汽车运行状态的感知、规划和控制，至今如火如荼。**第三步：机器自驾驶，自纠错，自成长。它比经验驾驶员更安全。**2000 年以来，我们摆脱了纯自动化的技术路线，强调用“机器驾驶脑”代替驾驶员开车，让车辆成为一个有感知、有认知、有行为、可交互、会学习、自成长的轮式机器人，成为移动的认知平台。**学习和记忆互相依赖，互相促进，已有的记忆是学习的基础，新的记忆又是学习的结果，反映了记忆的重塑和驾驶认知的自成长过程。**驾驶脑的结构图如图 1 所示。安装有驾驶脑的载体或平台的结构常常各不相同，如陆地行驶的车辆、空中飞行的无人机、水上航行的无人舰艇，它们的具身，包括机械、电气及结构差异大，各不相同，导致动力学与运动学特性各不相同，不同驾驶平台的传感器类型、特性、数量、安装位置也各不相同，导致传感器处理模块各不相同，由此构成的智能驾驶系统模块的数量、接口各不相同。**作为移动认知平台的无人驾驶车中，异构芯片数量少则上百个，多则上千甚至几千个，异构芯片种类也从 40 种上升至 150 多种，它们都是物质硬构体。**

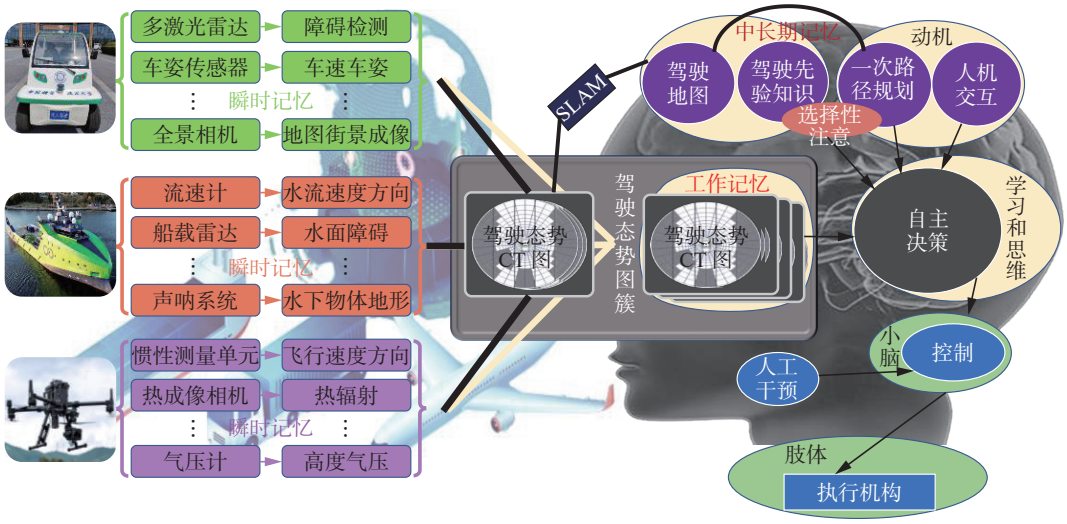

图 1　机器驾驶脑结构

Fig. 1　Machine driving brain structure

驾驶脑可分为三大组成块：物理空间的传感器信息处理模块，完成跨模态感知融合，**特别是雷达、相机和车辆所在位置地图的融合；**认知空间的思维决策模块，完成驾驶态势认知并形成决策，**特别要关注车辆实时拥有的路权；**物理空间对机器具身的运动控制模块，通过对车底盘的控制，给出对方向盘转角、油门（电机的转速和扭矩）和制动的控制量。和瞬时记忆相关的定位传感器，**特别是北斗等空间定位设备，要求能够达到厘米级导航**；车姿传感器包括感知车身的加速度和速度；视觉传感器看图像、看语义，雷达传感器看距离、看路权。对这些信息进行**跨模态的交互融合**，形成当前的驾驶态势图，送入工作记忆。在长期记忆里，有驾驶地图、交通规则、各类典型场景记忆棒和事故记忆棒等。除此以外，还要有人机交互，完成路径规划，要通过学习、思维完成自主决策，最后通过汽车的控制平台中的交互总线、决策总线和控制总线来完成汽车具身的运动学和动力学行为。驾驶脑中的思维决策模块，可以在不同传感器配置和各类异构平台上方便地移植。**感知、思维、决策、控制形成反馈回路，构成再感知、再决策、再控制的认知循环。**

2.2　基于智痕元胞的瞬时记忆、工作记忆和长期记忆网络

从机器驾驶脑这个典型案例可以看出，**记忆是真实世界不同尺度的摘要或者抽象，而非复本，也不是无损压缩。回溯过往是一种认知的逆时间旅行，得以摆脱时间总是向着未来的束缚，能够将过去和现在连接到一起，为认知提供了连续性，成为机器当前认知能力的基础。**它由瞬时感知记忆组块、短时工作记忆组块、长期经验记忆组块构成，瞬时感知记忆组块和长期经验记忆组块中的深度神经元胞网络，其结构相似，最大的区别是反应时间不同。组块之间协同工作，**确保记忆的形成、调控和提取，**其工作原理和结构关系见图 2。

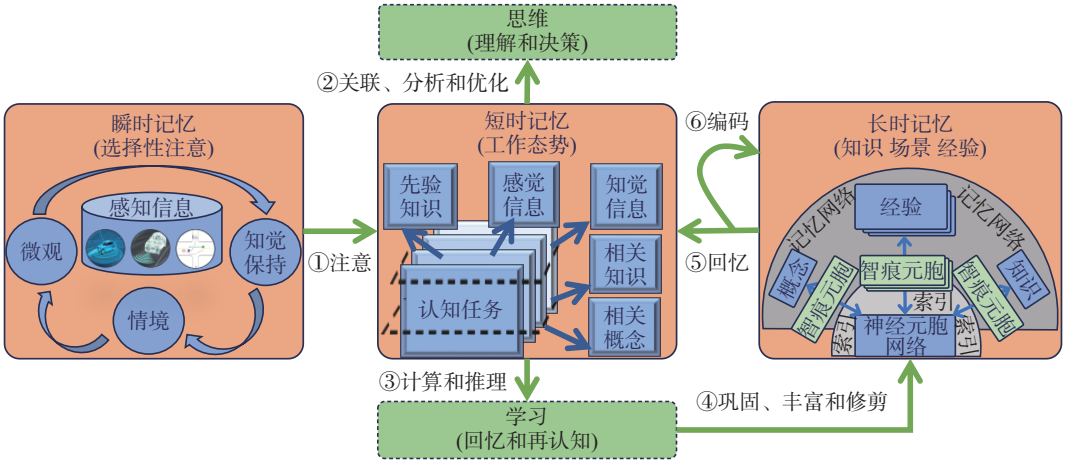

图 2　认知机器中记忆的形成、调控和提取

Fig. 2　Memory formation, regulation and retrieval in cognitive machines

瞬时记忆组块中的**智痕元胞网络，**完成对感知觉信息的存储，其记忆时间短，信息量丰富。选择性注意机制将抽取与当前任务相关的内容，实现感知觉保持，传递给短时工作记忆。短时记忆组块是根据认知任务对当前态势分析的结果，这时，要从长期的、已有的记忆中提取相关概念、知识，感知觉信息和长期形成的先验知识关联，分析、优化，以形成决策。当然，通过一次又一次的再认知，也会对长期记忆组块中的知识、场景与经验进行些许修剪，实现认知的自成长。在**短时工作记忆组块**区，包含大量的计算和推理，通过理解与决策实现机器具身智能的控制方案，最终反映到物理空间具身行为的操作。长时记忆区组块中的**智痕元胞网络，**积累了短时记忆经过复用、抽象、编码建立的丰富而牢固的、重要的概念、经验与知识，记忆时间长。智痕元胞节点与节点相互连接的边的强度和变化，模拟突触的可塑性，可以单向也可以双向。机器的智痕元胞网络虽然从初始认知核带来的元胞的总数量没有太大的变化，但随着认知的自成长可以重塑，即自我组织能力。有的智痕元胞逐渐对外有太多的连接，是记忆网络中的**重要路由节点。**而更多智痕元胞对外只有不太多的连接，是记忆网络中的**平凡路由节点。**更多的节点表现为网络的**边缘节点。**还有些节点在网络拓扑中体现骑墙，介于两个或者多个抱团节点集群之间，可称为**骑墙节点。**总体上，边的连接度大致服从幂律分布。相互连接的智痕元胞集群通过外界刺激或者通过学

习被激活，重要节点和平凡节点的区分不是一成不变的，而是由外界刺激动态激活的。记忆提取的是智痕元胞整体水平状态，是记忆网络的整体具象，有时甚至是涌现。拓扑状态数几乎是无限的，因此概念或知识的表达几乎是无限的。认知机器中 PB 级规模的神经元胞网络，如同互联网搜索用的网络架构，被称为人工神经元胞网络。**用智痕元胞模拟神经元记忆的残痕，用智痕元胞之间的连接强度和变化模拟突触的作用，用大规模人工神经元胞网络构成记忆网络。**记忆网络在学习和进化过程中不断地**微重构**，完成记忆的调整与控制，用当前记忆网络在不同侧面、不同尺度上的整体拓扑状态，表达并实现当前注意力对过去记忆的提取。在长期记忆中的许多智痕元胞常处于休眠状态，某一时刻被觉醒的那一小小部分是依靠注意力——当前待解决的问题来激活的，带有不确定性，**长期记忆的提取方式是再认知和回忆，这是与长时记忆中被唤醒的一个非常局部的网络的联想搜索问题。记忆中可以有相互冲突的知识，不断拉扯，形成决策。**人工智痕元胞网络构造清楚了，机器的记忆智能也就基本清楚了。**它随着认知的自成长，不断地修剪权重，改变连接，成为一个具有小世界特征的、无标度的复杂网络，具有抱团特征、层次结构、自相似性和长尾分布，可以用它来实现当前认知中不同尺度抽象记忆的形成、调整、控制和提取，成为记忆智能。记忆智能确保了多元认知，成为认知机器中多领域、多情境中计算智能的边界和约束。**它是非单调的、进化发展的，在不同时刻、不同情境会有不同应对，不完全收敛，不完全自洽，不整体统一。

在认知的机器中，深度学习是基于记忆的经验认知模式。它打破了算法被困在程序里面的窘境，开辟了用数据生成或者修改算法中的参数，生成知识的一条道路，成为人工智能历史上的一个新的里程碑。**但是深度学习不应该仅仅是预训练、预编程，最好不把训练和使用截然分开，要在实时的交互和迭代中进行深度学习，边指导、边学习、边使用，自学习，自纠错，微调记忆。**概念由学习而产生，是新形成的记忆关联的基础，知识则是概念和概念之间的关系。记忆留存在大脑中被记忆网络动态重构并使用，这样的记忆网络是叠层、多侧面、多尺度的复杂网络，可能抽象成语言网络、情境网络、程序网络，语言网络又有多个侧面，如语义网络、语用网络、语境或者语法网络。通过这些网络在不同尺度上的表现和状态，生成概念、泛概念树或者知识图谱，其中概念或者泛概念树是支持记忆认知的关键组成部分。它通过随后在学习经历中出现的刺激而被一次次激活，**所有这些同时发生的激活形成了相互连接的记忆网络的整体具象，构成一次次记忆知识的提取。**

不同的记忆认知任务，使用不同的标记和索引方法，最终可以在多个脑的记忆区得到相互印证的结果。**首先，**标记的智痕元胞仅通过相应的条件刺激而被激活，而不会被与训练经历无关的刺激而激活。**其次，**仅仅智痕元胞集群的一小部分的再活化，理论上会导致整个集群的再活化。**此外，**如果因为遗忘而淡化了一小部分智痕元胞或集群元胞，并不一定会导致整个记忆表征的破坏，可能只是适当地降低了记忆能力而已。相反，如果因为相同情境的频繁感知和刺激，有可能会加深印迹，或者导致新的连接。**还有，**在没有感知刺激的情况下，通过经验的回忆也可激活智痕元胞诱导记忆的提取，成为常识或者本能。**最后，**如果激活机制缺失，即使智痕元胞存在，也会阻挡随后的记忆提取，表现为记忆丢失。

2.3　机器具身行为的多重嵌套控制

机器具身智能是所有人工智能研究的出发点和归宿。机器具身行为是思维、决策与控制的结果。认知机器中多重嵌套的交互回路设计确保机器具身智能的稳定性和可成长性，完成机器的使命，机器具身行为控制原理如图 3 所示。这里有三重嵌套回路：**最外层**是当前位置环境变化的交互回路，体现任务使命的对齐；**中间层**是当前注意力选择的交互回路，聚焦实时态势；**最内层**是具身行为的运动学与动力学的交互回路，体现自动反馈。三重嵌套控制涉及认知空间与物理空间。在认知空间中，进行情境感知、跨模态融合，形成瞬时记忆，在工作记忆中，它们通过当前态势的“判断黑板”，在记忆约束下进行计算，进而在当前环境下进行态势判断和推理，产生决策，同时在长期记忆里进行记忆提取，用过去的经验反馈当前态势下的应对，用注意力选择来聚焦当前态势中的主要矛盾，用再认知的结果对知识、场景与经验进行修饰，微调长期记忆。在物理空间中，它们实现运动学与动力学控制，完成机器具身行为动作，再由运动姿态传感器进行状态反馈。通过多重嵌套的交互控制回路，认知机器逐渐地理解人设定的任务目标，完成使命。

智能机器与外部世界发生交互，在物理环境里做出行为动作，并接收机器承载的传感器感知

得到的反馈。感知和行为是机器具身智能这枚“硬币”的两个面。这种交互是让机器获得认知动机和意图，感知和预测外部世界时的正常途径，也是强化学习、深度学习中正常的感知方式。互动才能与外部世界沟通，例如奖励代表着人的意图和目标，智能机器总是希望能够获得最大化奖励。实际上，**认知机器始终都是在和环境的连续交互中体现认知行为的试探和反馈，**交互认知不仅是冯·诺依曼架构中的输入和输出，也不仅是友好用户界面设计、图形交互界面设计、拟人化交互服务的方法学问题，它更是机器认知的自学习、自纠错、自成长、知行合一、人机共生的问题。人工智能过去十年中的标志性成就——深度学习——要求海量的人工标注就可以看作是一种交互的过程。**认知机器的最基本特征是能够在与环境的交互过程中学习和成长。**无论是指导学习（Supervised Leaning，有人把它译为监督学习，不很确切），或者半指导学习，或者自主学习，都需要不停地交互。认知机器的组成中，可能有定时定位定姿、语音文字、图形图像等多类子部件，如各种雷达、摄像头、语音文字通信、卫星定位接收机等感知设备。**连接思维和物理世界最好的桥梁就是人体或者机器具身。**认知机器的具身，可能是车辆、飞机、船舶、盾构机、多自由度机器人等，承担与外界环境的感知和交互。交互感知是双方无法单独产生的，感知中有试探，有模仿，有反馈，是不确定性认知的重要来源，是机器进行决策的重要前提和实现手段。

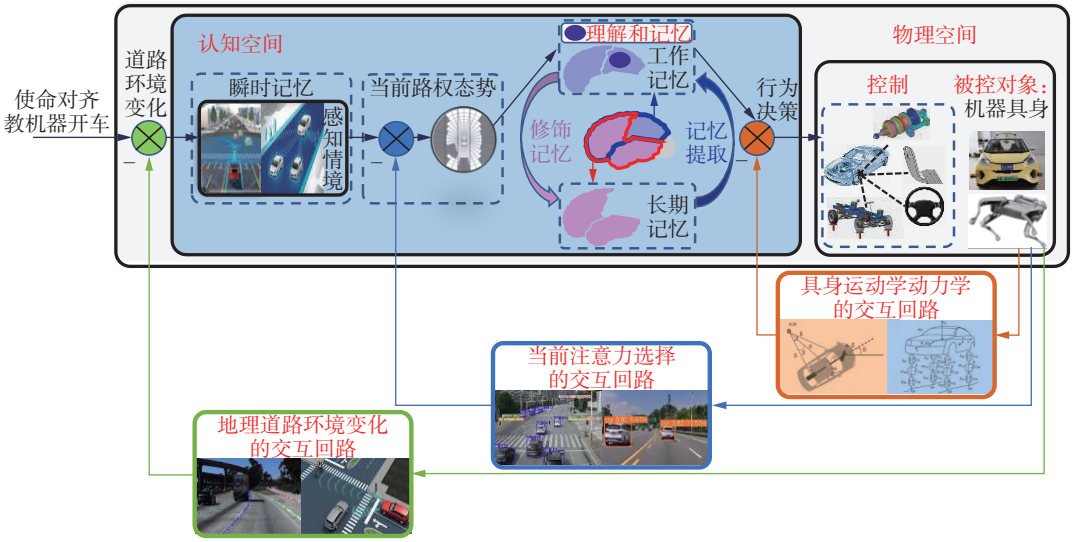

图 3　机器具身行为的多重控制原理

Fig. 3　Multiple controls of embodied machine behavior

2.4　可交互、会学习、自成长的认知机器的通用架构

认知机器已经越过了算力、算法和数据 3 个硬核的阶段。机器中的瞬时记忆组块和短时记忆组块，除了 CPU，还可根据需求采用 DPU、GPU、TPU、FPGA、SSD、搜索引擎等并行处理单元，而计算组块则可采用 CPU 和 GPU 等处理器实现，也有可能采用处理效率更高的 3D 存算一体化。**有的组块里，DPU 为核心，CPU 围绕 DPU 转；有的组块里，GPU 为核心，CPU 围绕 GPU 转。总之，新架构中的系统架构一定是分布式的、并行的、异构的，甚至是超异构的，只要它们能够和机器的时序整体上合拍，能实时地进行数据交互即可，认知机器的时间精度越高，并行效率越高。**

可交互、会学习、自成长这三方面成为认知机器的新硬核，其最基本的特征是能够在与环境的交互过程中学习、纠错和成长，可以接受指导学习和强化学习，也可以自主学习，增强记忆。认知机器的学习和作业，包括先入为主、赋予任务、引导、释疑、解惑、交互认知、监督等有指导的学习。自主学习是把指导学习的结果转为长期记忆的重要环节，例如复习、消化理解、自己纠错。如果简单地把指导学习称为有监督学习，自主学习称为无监督学习，就过于简单化了。一次性学习之后常是短期记忆，间隔性地重复学习有利于形成和巩固长期记忆，重复学习的时间间隔非常重要，充满不确定性，体现自纠错和长期记忆的自成长能力。ChatGPT 在训练过程中高薪聘请了“提示工程师”。同理，在认知机器中也需要“指导工程师”。人与机器能有效沟通完成预设任务，人教机器学，机器自主学，机器逐渐地理解人设定的任务目标，其统一的过程可称为使命对齐，精准完成作业，具身体现智能。机器会学习包括 3 个环节：专家操作，机器学习；机器自动运

行，人干预；机器自操控、自学习、自纠错、自成长。**这 3 个环节循环迭代，实现有指导学习、半/弱指导学习、自主学习。**可交互、会学习、自成长的认知机器的通用架构如图 4 所示。

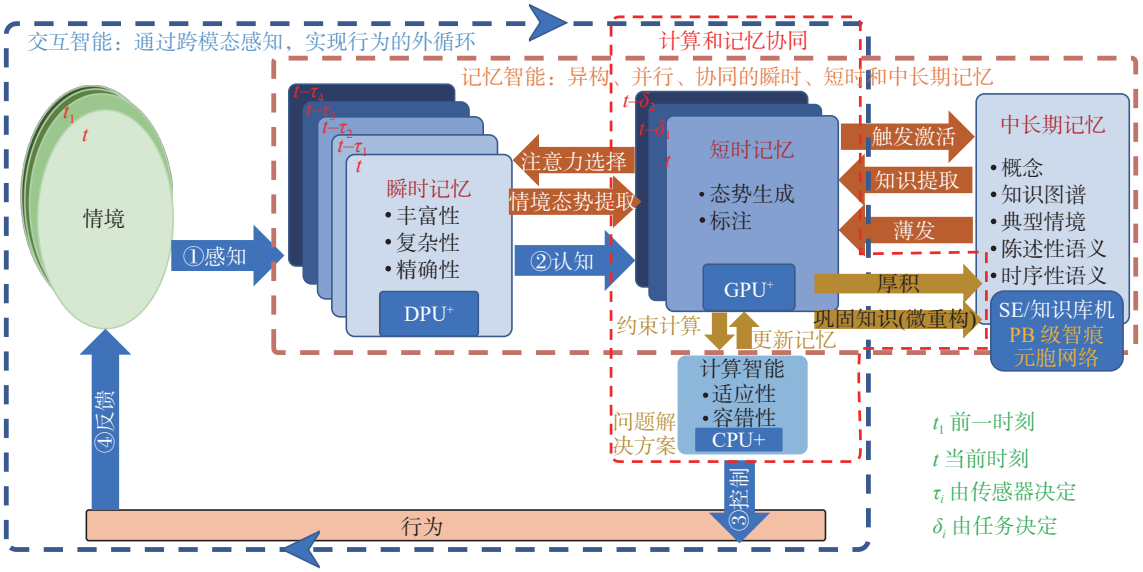

图 4　认知机器的通用架构

Fig. 4　A general architecture of cognitive machines

注意力机制触发长期记忆，进行相关知识提取；短时记忆通过学习，形成并巩固记忆智能；记忆与计算协同，记忆约束计算，计算更新记忆，并利用计算智能提高记忆智能的可解释性。

记忆、交互和计算三部分之间的信息传递机制复杂并螺旋发展。每个记忆系统依然嵌有交互与计算子系统，但整体表现为记忆。认知机器通用架构可在结构上体现人脑不同记忆区不同记忆留存的认知网络，体现已有的人工神经元胞记忆网络节点中的印迹、概念、概念树、知识图谱中的不确定性和多索引机制，体现情境驱动的记忆智能，实现了情境数据和知识模型双驱动。而计算智能作为认知机器通用架构中的一个重要组成，多种计算模型可以分别是已有的记忆对当前计算的多元约束和指导监督。

3　机器中的认知核和宕机后的再激活

3.1　激活机器的钥匙：时钟、时序和递归

机器中的软构体是承载或者寄生在硬构体上的，如同人的精神寄生在硬构体之上一样。当然，它也可以寄生在已有的其他软构体上。机器里一定要有一个最基本的时钟，而时钟赖能量为生，时间寄生在时钟上，形成时序。**激活机器的钥匙是时钟、时序和递归。**认知核中的物质硬构体和思维软构体在加电后的纠缠，表现在时钟、芯片、机器主板、BIOS（basic input output system）和 OS 在自举状态的递归复用，才让机器“活”起来。作为工具的机器，结构寄生在物质上。要激活机器，需要能量，能量激活时钟，时钟产生节律。如同生命有节律一样，机器利用时钟形成时间和节律，可以在当前的周期内为下一个周期提供一个更新的输入，**总是存在下一个周期能够保持思维的连续性，机器思维才能活动起来。创造机器智能这样的人造物扩展人类智能，这是图灵的划时代贡献，堪与牛顿、爱因斯坦媲美，可惜很多人对此认识不足。**正是图灵和冯·诺依曼的计算机体系结构设计中的 CPU，保证了指令和数据一样存储，指令和数据形式上并无区别。将程序指令存储器和数据存储器合并在一起，顺序执行程序，让机器能够自举。依靠只读存储器中的基本输入输出系统（ROM-BIOS）引导。基本输入输出系统 BIOS 是一组固化到只读存储器 ROM 芯片上的程序。**在 BIOS 引导下，机器启动时加载的第一批控制指令，所有后续的物质硬构体和思维软构体，类似于承载生命基因编码的 DNA，被称之为机器初始的认知核。**这个只读存储器是把结构和时间完全寄生到物质和能量上的客观存在，规定了机器基本的输入输出次序，包括开机后自检程序和系统自启动程序，为机器提供最底层的、最直接的硬件设置和控制，体现了硬构体和软构体之间的纠缠，然后激活操作系统。**整个过程是认知核中的硬构体和软构体纠缠的正反馈过程，导致涌现。**物质、能量、结构和时间之间的这种纠缠状态，可类比为“薛定谔的猫”，导致新

的宏观有序状态, 认知就绪, 机器从原先的“死物质”变“活”了。

3.2　宕机后的再激活

生命不能重来, 机器可以关闭后重启。认知核包含机器具身物质硬构体, 如时钟、集成电路芯片、主板等, 也包含思维软构体, 如机器指令、BIOS 和 OS 等。机器如果没有了能量供给, 如断电, 便会停止工作; 恢复供电后机器又可以再次自举, 通过激活操作系统, 重新进入认知的工作状态。但是, 硅基机器中的物质硬构体不能自繁衍、自成长、自修复, 只能被组装、被生产、被修复。硬构体老化了、失灵了, 被修复之后可以重启, 死活多次。如果有新的硬构体、软构体加进来, 只要适配, 升级之后, 可以提高机器认知的性能。**硅基机器可通过认知核更新, 完成升级换代。**

4　总结和展望

针对“生命是什么?”“信息是什么?”“智能是什么?”“机器如何认知?”等当前全人类最为关注的基础科学问题。本文在人类完全弄清楚非生命和生命之间的双向转变机理之前, 试图能够弄清楚机器如何思维? 如何像人又不像人? 这就找到了从生命转变到非生命的一个重要突破口, 就找到了从碳基生命智能转变到硅基机器智能的物理通道。

对于人来说, **思维具有连续性、随机性和模糊性, 缺少形式上的精确性**。如果将“翻译”和“查重”这样的认知任务扩大到检查全球范围的学术论文抄袭, 这几乎是不可能完成的事, 但对于机器而言却无太多困难。基于时序的机器认知, 可以做到纳秒、皮秒、甚至飞秒级的反应速度。倘若机器以飞秒计算, 人以秒计算, 一飞秒与一秒的比例是 $1:10^{15}$, 相当于一秒和 3 200 万年的比例。机器可以用皮秒和飞秒模拟人的思维活动周期。思维软构体通过抽象和联想, 自我复制, 自我拓展, 引发类比, 使得认知具有一般性和普遍性。**机器可以用足够多逼近无穷多, 用足够大逼近无穷大, 用足够小逼近无穷小, 用足够精确量替代连续量, 模拟重演物理、化学、生物、材料等学科中大多数快速变化的过程,** 例如, 通过虚拟现实展示单分子的振动和转动、化学键的断裂和形成等, 可见机器暴力思维的威力。机器能通过暴力计算和暴力仿真, 完成蛋白质折叠的三维结构预测, 并不奇怪。**物质硬构体很难约束想象的范围和思维的内容, 思维如果不和物理世界实时沟通验证, 处理得不好, 机器也可能过度幻想, 陷入思维的死循环, 难以自拔。**

机器学习的结果是去微调机器里的长期记忆, 即微调人工智痕元胞的网络拓扑, 实现认知的自成长。可喜的是, 认知机器可以大批量复制, 而且机器自身又可以持续学习。不久的将来, 人与机器交互, 人教机器, 机器教人, 协同创新。认知机器会越来越多地发明出新材料的配方, 编写出最新的乐曲, 创作出更美的图片和视频, 自动证明新的数学猜想, 制造打印各种复杂 3D 结构的形状, 提出新的学科假设, 驱动产生新的科学发现。**认知机器成为人类思维的超强加速器和智能行为的超强放大器, 人机交互协同创新, 机器可以和科学家、工程师一同做出发现、发明和创造。至于是不是机器做出的创造, 已经不再重要。人类需要认知机器, 认知机器需要人类, 互相激励。人类依然是主宰, 机器越智能, 人类也越智慧。**

参考文献:

[1] 李德毅. 论智能的困扰和释放[J]. 智能系统学报, 2024, 19(1): 249−257.
LI Deyi. On the puzzle and release of intelligence[J]. CAAI transactions on intelligent systems, 2024, 19(1): 249−257.

[2] TURING A M. Computing machinery and intelligence[M]. Netherlands: Springer, 2009.

[3] 李德毅, 刘玉超, 任璐. 人工智能看智慧[J]. 科学与社会, 2023, 13(4): 131−149.
LI Deyi, LIU Yuchao, REN Lu. Viewing wisdom from the perspective of artificial intelligence[J]. Science and society, 2023, 13(4): 131−149.

[4] 赵南元. 认知科学与广义进化论[M]. 北京: 清华大学出版社, 1994.

[5] 桑基韬, 于剑. 从 ChatGPT 看 AI 未来趋势和挑战[J]. 计算机研究与发展, 2023, 60(6): 1191−1201.
SANG Jitao, YU Jian. ChatGPT: A glimpse into AI's future[J]. Journal of computer research and development, 2023, 60(6): 1191−1201.

[6] 李德毅, 郑思仪, 黄立威, 等. 认知的形式化[J]. 中国基础科学, 2024, 26(2): 1−14.
LI Deyi, ZHENG Siyi, HUANG Liwei, et al. The formalization of cognition[J]. China basic science, 2024, 26(2): 1−14.

[7] 李德毅. 人工智能基础问题: 机器能思维吗?[J]. 智能系统学报, 2022, 17(4): 856−858.
LI Deyi. Artificial intelligence fundamental question: Can machines think ?[J]. CAAI transactions on intelligent sys-

tems, 2022, 17(4): 856−858.

[8] LI Deyi. Cognitive physics: the enlightenment by schrödinger, turing, and wiener and beyond[J]. Intelligent computing, 2023, 2: 0009.

[9] FORD L, LING E, KANDEL E R, et al. CPEB3 inhibits translation of mRNA targets by localizing them to P bodies[J]. Proceedings of the national academy of sciences, 2019, 116(36): 18078−18087.

[10] PENN A C, ZHANG C L, GEORGES F, et al. Hippocampal LTP and contextual learning require surface diffusion of AMPA receptors[J]. Nature, 2017, 549(7672): 384−388.

[11] ASSRAN M, DUVAL Q, MISRA I, et al. Self-supervised learning from images with a joint-embedding predictive architecture[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 15619−15629.

[12] GUPTA A, YU L, SOHN K, et al. Photorealistic video generation with diffusion models[EB/OL].(2023−12−11)[2024−09−20]. http://arxiv.org/abs/2312.06662.

[13] 李德毅, 殷嘉伦, 张天雷, 等. 机器认知四要素说[J]. 中国基础科学, 2023, 25(3): 1−10,22.
LI Deyi, YIN Jialun, ZHANG Tianlei, et al. Four most basic elements in machine cognition[J]. China basic science, 2023, 25(3): 1−10,22.

[14] URMSON C, ANHALT J, BAGNELL D, et al. Autonomous driving in urban environments: Boss and the urban challenge[J]. Journal of field robotics, 2008, 25(8): 425−466.

[15] ZHANG Xinyu, GAO Hongbo, GUO Mu, et al. A study on key technologies of unmanned driving[J]. CAAI transactions on intelligence technology, 2016, 1(1): 4−13.

[16] 李德毅, 马楠, 高跃. 未来汽车: 会学习的轮式机器人[J]. 中国科学: 信息科学, 2020, 63(9): 255−262.
LI Deyi, MA Nan, GAO Yue. Future vehicles: learnable wheeled robots[J]. Science China information sciences, 2020, 63(9): 255−262.

[17] HU Yihan, YANG Jiazhi, CHEN Li, et al. Planning-oriented autonomous driving[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver: IEEE, 2023: 17853−17862.

[18] 李德毅, 高洪波. 基于驾驶脑的智能驾驶车辆硬件平台框架[J]. 工程(英文), 2018, 4(4): 464−470.
LI Deyi, GAO Hongbo. A hardware platform framework for an intelligent vehicle based on a driving brain[J]. Engineering, 2018, 4(4): 464−470.

[19] ENGLE R W, TUHOLSKI S W, LAUGHLIN J E, et al. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach[J]. Journal of experimental psychology: General, 1999, 128(3): 309.

[20] 李德毅. 脑认知的形式化: 从研发机器驾驶脑谈开去[J]. 科技导报, 2015, 33(24): 125−125.
LI Deyi. Formalization of brain cognition: talking from the development of robot driving brain[J]. Science & technology review, 2015, 33(24): 125−125.

[21] LECHNER M, HASANI R, AMINI A, et al. Neural circuit policies enabling auditable autonomy[J]. Nature machine intelligence, 2020, 2(10): 642−652.

[22] 林龙年. 发现大脑定位系统的细胞组构[J]. 科学, 2015, 67(1): 30−34.
LIN Longnian. The discovery of cells constituting brain's GPS[J]. Science, 2015, 67(1): 30−34.

[23] PAN Feng, BAO Hong. Preceding vehicle following algorithm with human driving characteristics[J]. Part D: Journal of automobile engineering, 2021, 235(7): 1825−1834.

[24] LIANG Tianjiao, PAN Weiguo, BAO Hong, et al. Bird's eye view semantic segmentation based on improved transformer for automatic annotation[J]. KSII transactions on internet and information systems, 2023, 17(8): 1996−2015.

[25] LIU Kang, ZHENG Ying, YANG Junyi, et al. Chinese traffic police gesture recognition based on graph convolutional network in natural scene[J]. Applied sciences, 2021, 11(24): 1−19.

[26] 马楠, 高跃, 李佳洪, 等. 自驾驶中的交互认知[J]. 中国科学: 信息科学, 2018, 48(8): 1083−1096.
MA Nan, GAO Yue, LI Jiahong, et al. Interactive cognition in self-driving[J]. Science China information sciences, 2018, 48(8): 1083−1096.

[27] WIENER N. Cybernetics: or control and communication in the animal and the machine[M]. Massachusetts: MIT Press, 1948.

[28] 李德毅. 机器的交互式具身智能[J]. CAAI 人工智能研究, 2024, 3: 1−6.
LI Deyi. Interactive embodied intelligence of machines[J]. CAAI artificial intelligence research, 2024, 3: 1−6.

[29] FUSTER J M. Distributed memory for both short and long term[J]. Neurobiology of learning and memory, 1998, 70(1/2): 268−274.

[30] ZHENG Ying, HONG Bao, MENG Chaochao, et al. A method of traffic police detection based on attention mechanism in natural scene[J]. Neurocomputing, 2021, 458: 592−601.

[31] XIA Jing, CHEN Nanguang, QIU Anqi. Multi-level and joint attention networks on brain functional connectivity for cross-cognitive prediction[J]. Medical image analysis, 2023, 90: 102921.

[32] 李德毅. 人工智能看哲学[J]. 科学与社会, 2023, 13(2): 123−135.
LI Deyi. Artificial intelligence views philosophy[J]. Science and society, 2023, 13(2): 123−135.

[33] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115−133.

[34] 李德毅. 新一代人工智能十问[J]. 智能系统学报, 2020, 15(1): 1.
LI Deyi. Ten questions for the new generation of artificial intelligence[J]. CAAI transactions on intelligent systems, 2020, 15(1): 1.

[35] 李德毅. 机器如何像人一样认知: 机器的生命观[J]. 中国计算机学会通讯, 2022, 18(10): 11−14.
LI Deyi. How machines cognize like humans: the machine's view of life[J]. Communications of CCF, 2022, 18(10): 11−14.

[36] 李德毅, 马楠. 人工智能看教育[J]. 高等工程教育研究, 2023, 3(5): 1−7.
LI Deyi, MA Nan. Viewing education from the perspective of AI[J]. Research in higher education of engineering, 2023, 3(5): 1−7.

[37] NEUMANN J, BURKS A W. Theory of self-reproducing automata[M]. Urbana: University of Illinois press, 1966.

作者简介:

李德毅, 清华大学博士生导师, 中国工程院院士, 欧亚科学院院士。中国人工智能学会和中国指挥控制学会名誉理事长。主要研究方向为不确定性人工智能、数据挖掘、复杂网络、自驾驶和认知物理学。E-mail: lidy@cae.cn。