

大脑建模的理论框架及热点问题

胡东伟¹, 冯晓璐²

(1. 中国电子科技集团公司第五十四研究所, 河北 石家庄 050080;

2. 石家庄二中实验学校, 河北 石家庄 050080)

摘要: 理解大脑的工作原理是实现“终极”人工智能的重要途径。建模并验证模型的正确性是理解大脑工作原理的重要手段。首先概述大脑的实验研究方法和建模研究方法; 然后给出大脑的框图模型, 并以强化学习为核心介绍大脑建模的理论框架; 最后给出目前亟待解决的几个热点问题, 并讨论大脑建模与相关学科之间的关系。

关键词: 大脑; 模型; 人工智能; 机器学习; 前额皮层; 海马; 杏仁核

中图分类号: TP18

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202141

Theoretical framework of brain modelling and highlighted problems

HU Dongwei¹, FENG Xiaolu²

1. The 54th Institute of China Electronics Technology Group Corporation, Shijiazhuang 050080, China

2. The Shiyuan School Attached to Shijiazhuang No.2 Middle School, Shijiazhuang 050080, China

Abstract: Understanding how the brain works is an important way to constructing the final artificial general intelligence. Building models, and testing the correctness of these models play a pivotal role on understanding how the brain works. Firstly, the experimental techniques and mathematical modelling methods for the study of brain were reviewed, and then the block diagram model and theoretical framework were shown, with an emphasis on reinforcement learning. Finally, some highlighted problems in this area, and the relation with closely related disciplines were addressed.

Key words: brain, model, artificial intelligence, machine learning, prefrontal cortex, hippocampus, amygdale

1 引言

近年来, 由于深度学习^[1]的发展, 人工智能, 特别是通用人工智能 (artificial general intelligence, AGI) 研究在世界范围内掀起了一股新的热潮。然而, 深度学习毕竟只是人工智能领域的一个方面, 不足以上升到人工智能全貌的高度。回顾人工智能领域 60 余年的发展可以看到, 关于通用人工智能的基本框架和理论体系, 仍然莫衷一是。在这方面, 我国学者也做了很多的努力和探讨, 提出了机制主义^[2]、泛逻辑学^[3]等概念^[4-6], 试图从更高的视野、更基础的理论, 建立和统一人工智能理论体系。

智能的源泉在于人类大脑。倘若能够搞清楚人

类大脑的工作机理, 人工智能的实现就会水到渠成, 甚至有望达到其“终极”形态。这在人工智能界早有共识^[7-8]。然而, 困难又转移到了理解大脑的工作机理上。幸运的是, 近年来生物神经科学飞速发展, 为研究者理解大脑的工作机理提供了很多线索。因此, 从大脑的计算建模出发, 不失为迈向通用人工智能的一条途径。此外, 理解大脑的工作机理, 不仅将促进人工智能的发展, 同时将有治疗瘫痪、癫痫、帕金森病、阿尔茨海默病等神经系统疾病^[9], 直接服务于人本身, 这将具有更紧迫的现实意义。从这个角度来说, 建立大脑的计算模型, 理解大脑的工作机理, 超越了人工智能这门技术科学的范畴, 达到了自然科学的高度。

收稿日期: 2020-07-14; 修回日期: 2020-08-14

通信作者: 胡东伟, hudw1980@sina.com

本文试图从“大脑建模”这一角度出发,给出大脑理论模型的自顶向下的全貌。在这个全貌之下,可以统一“人工智能”“机器学习”知识体系的诸多方面,弥补其“浅层化”“碎片化”的不足^[2]。而且,“大脑建模”这一主题,也将为人工智能、机器学习理论提出新的难题,指出新的理论研究方向。此外,本文也期望引导一些从事人工智能、机器学习、大数据等应用研究的工程师们进入大脑研究领域,成长为科学家。

基于上述目的,本文的重点将在大脑的模型架构和核心功能——大脑意识的建模研究上,而对大脑的输入输出系统不做过多介绍,例如视觉系统、听觉系统以及脊髓等;与大脑无关或联系不紧密的机器学习算法(例如卷积神经网络(convolutional neural network, CNN)^[10])亦不会提及。本文的目的是给出大脑工作机理的理论框架。

本文结构如下:第2节简述大脑的基本知识,第3节简单介绍大脑的实验研究方法,第4节介绍常用的大脑的建模研究方法,第5节描述大脑的框图模型,第6节描述大脑的数学模型,第7节给出大脑建模研究面临的热点问题,并讨论大脑建模与机器人研究、数据科学研究之间的关系,第8节总结全文。

2 大脑的基本知识

大脑是人体的一个重要器官。它是人思想意识的物理依托,是人的感觉、意识、情感、记忆、动机、目标的贮藏地,是人行行为/行动的总指挥。人的一切智能活动都由大脑来主导和控制^[11-12]。

现代生物研究表明,人类的一切智能活动依托于人的神经系统,而大脑为神经系统的最高级部分。它由左、右两个半球构成。大脑半球表面的一层灰质被称为大脑皮层。大脑皮层是神经细胞(也称神经元)的集中地。大脑内含有约860亿个神经细胞^[11-12]。神经细胞和神经细胞之间由突触连接。每个神经细胞大约与1 000~10 000个突触相连,如此构成一个巨大的神经网络。正是这个神经网络构成了人思维的“机器”,并主导人体内的一切活动,调节人的身体与周围环境的平衡。

信息以神经脉冲的形式由一个神经细胞传向另一个神经细胞。神经细胞兴奋时就会释放神经脉冲,并通过它的轴突传递给后续神经细胞的树突。因此,脑内工作的神经网络是一种“脉冲神经网

络”。根据神经细胞释放的神经脉冲是会激励后续神经细胞的兴奋,还是会抑制后续神经细胞的兴奋,可将神经细胞分类为兴奋性神经细胞(excitatory neurons)和抑制性神经细胞(inhibitory neurons)^[13]。

现代神经科学研究表明,大脑的功能是分区的。例如视觉皮层(V1~V5)位于大脑后部,主要负责视觉信息的处理;运动皮层(motor cortex)位于大脑顶部,是大脑中参与计划、控制和执行自主运动的区域;前额皮层(prefrontal cortex, PFC)是大脑中负责思考和决策的部分,它是大脑的首席执行官;人的情感主要由杏仁核(amygdala)来控制;而空间记忆与大脑的海马(hippocampus)密不可分^[12,14]。现代生物研究表明,虽然脑内各处的神经细胞和突触形状各异,但其基本结构和化学成分几乎相同。因此,大脑皮层各区域的功能主要是由其网络连接(包括内部连接和边界条件)决定的。因此,解密大脑皮层的神经网络结构,理解这个神经网络的行为,被认为是理解大脑工作机理的一个重要方向。

3 大脑的实验研究方法

科学理论的建立需要从实验数据出发,并得到实验数据的证实,大脑的研究也一样。大脑研究的最初始数据从解剖学得来。著名的病例亨利·莫莱森(Henry Molaison)就是手术切除了海马组织以及海马周围的部分内侧颞叶组织,导致不能形成长期记忆,从而确认了海马区与长期记忆的形成之间的联系^[11-12]。但是,随着科学研究伦理的建立,解剖学的研究方法受到了极大的限制。现代的大脑研究更多采用系统辨识的方法。大脑的系统辨识原理如图1所示。在图1中,首先根据猜测提出一个大脑的模型,并给大脑和该模型同时施加激励,观察大脑和模型的输出,力求二者的输出一致。通过比较大脑和模型输出之间的误差,可进一步修改调整模型的结构和参数。

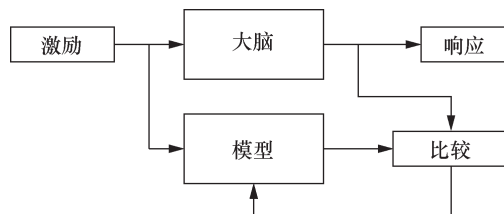


图1 大脑的系统辨识原理

3.1 大脑的激励方法

大脑的激励方法分为两种：感觉刺激方法和电极刺激方法。

3.1.1 感觉刺激方法

感觉刺激方法通过人体的感觉器官给大脑施加激励。例如，通过视觉给大脑施加激励，可以让人观察不同的图片^[15]；通过听觉给大脑施加激励，可以让人听不同的声音；通过嗅觉给大脑施加激励，可以让人闻不同气味的液体或气体等。还可以通过更复杂的、各种感觉器官的联合刺激，对大脑施加更复杂的激励。例如，要研究情绪对记忆的影响，就需要调控人的情绪，这可以通过在一个特定装饰的空间内，给人看特定情节的电影来实现。

感觉刺激方法的特点是对人体/大脑没有损害，因此实施起来较为容易；缺点是各种不同的激励因素混合在一起，难于解耦合。

3.1.2 电极刺激方法

电极刺激方法将一根或多根电极紧贴头皮或插入大脑皮层内，直接刺激大脑的神经系统。这类方法可以实现较精细的刺激，甚至可以对单个神经细胞进行刺激。这类刺激方法存在两个缺点。第一，存在分辨率不高的问题。如果每个电极插入大脑皮层内，只刺激单个神经细胞，则由于电极的分辨率有限，相邻神经细胞无法直接刺激；如果每个电极紧贴头皮，对一片大脑皮层进行刺激，则无法对相隔较近的神经细胞实现不同的刺激。第二，插入大脑皮层的电极会对大脑产生损伤^[16]。

电极刺激方法已被用于某些神经系统疾病（例如癫痫）的治疗。它对未来的大脑增强系统、神经义肢系统等的开发也具有重要意义。

3.2 大脑的响应观测方法

大脑的响应观测方法包括脑机接口（brain computer interface, BCI）方法、脑成像方法和行为观测方法。

3.2.1 脑机接口方法

脑机接口方法主要是指采集分析脑电波信号的方法，它分为无损的脑机接口和有损的脑机接口^[15]。无损的脑机接口通过紧贴头皮的多通道电极感应大脑皮层内神经细胞的兴奋状态。目前常用的有64通道、128通道、256通道电极的BCI设备。每个电极感应的是一小片区域内无数神经细胞兴奋放电状态的叠加，称之为局部场电位（local field potential, LFP），这是一个连续的模拟信号。当某

片区域内有大量的神经细胞兴奋放电时，LFP电压较高；当大量的神经细胞被抑制时，LFP电压较低。多通道同时记录脑电波信号可形成脑电图（electroencephalogram, EEG）。

有损的脑机接口通过探针将电极插入大脑皮层内部，可探测单个神经细胞的放电状态。此时，电极感应到的是单个神经细胞的膜电压，可观察到锯齿状的神经脉冲。

3.2.2 脑成像方法

脑成像方法有很多种，包括功能性磁共振成像（functional magnetic resonance imaging, fMRI）、脑磁图（magnetoencephalography, MEG）、单光子发射计算机断层成像（single-photon emission computerized tomography, SPECT）和正电子发射断层成像（positron emission tomography, PET）等，其中fMRI比较常用。fMRI由美国贝尔实验室的学者小川诚二（Seiji Ogawa）于1990年发明。fMRI通过MRI仪器检测处于激活状态的脑区域内血氧水平和血流变化获取数据。这种血氧变化被称为血氧水平依赖（blood oxygenation level dependent, BOLD）。局部脑区在激活状态下，脱氧血红蛋白下降，表现为fMRI信号增高。

脑成像方法大多为无损方法。无损的脑机接口和脑成像方法的时间/空间分辨率如图2所示^[16]。无损的脑机接口的时间分辨率较高，可达到毫秒级别，但空间分辨率有限；脑成像方法的空间分辨率较高，但时间分辨率较低。

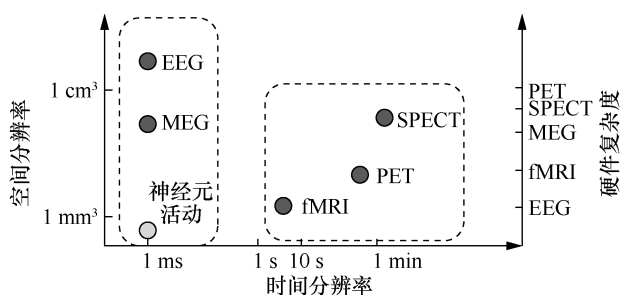


图2 无损的脑机接口和脑成像方法的时间/空间分辨率

3.2.3 行为观测方法

人所有的行为都是由大脑驱动的。通过给受试人安排某项任务并观察受试人的行为，可以分析大脑的工作机理。这是一种无损的方法，实施起来比较容易。但人的行为涉及的因素较多，难以把各个因素解耦合。若要解耦合，需要非常精巧的任务设计。

从复杂的联合感官刺激和行为观测可以看到，大脑的研究方法与心理学、行为科学紧密相关。

4 大脑的建模研究方法

如上文所述，大脑的模型由猜测出发，但要求模型的响应和大脑的真实响应尽量符合。本节给出目前常用的几种建模方法。

4.1 贝叶斯网络法

贝叶斯网络法是一种通用的机器学习建模方法。它以贝叶斯公式为基础，具有学习和推理的能力。贝叶斯公式如下：

$$p(A|D) = \frac{p(A)p(D|A)}{p(D)} \quad (1)$$

如式(1)所示，求解在观测到数据集 D 的条件下事件 A 发生概率可以分为学习和推理两个阶段。在学习阶段，给定数据集 D ，给定在数据集 D 条件下事件 A 发生的概率 $p(A|D)$ ，和事件 A 发生的先验概率 $p(A)$ ，可以推算出在事件 A 发生的条件下，数据集 D 的概率 $p(D|A)$ 。这里往往假定 $p(D|A)$ 符合某种分布，学习得到的往往是 $p(D|A)$ 分布的参数。在推理阶段，给定数据集 D ，根据 $p(D|A)$ 分布的参数可以计算出 $p(D|A)$ ，于是，根据事件 A 的先验概率即可推算出在观测到数据集 D 的条件下，事件 A 发生的概率。

以上示例只是简单的一个数据集和一个变量/事件的推理。当有较多变量时，各个变量之间的依赖关系就形成一个网络。该网络可以使用有向无环图来表示，这个网络就被称为贝叶斯网络。贝叶斯网络表示了变量/事件之间的因果关系。每个节点的入度表示了该变量/事件将受多少个其他变量/事件的影响；每个节点的出度表示了该变量/事件将影响其他多少个变量/事件。无入度的节点被称为尾节点；无出度的节点被称为头节点。尾节点表示了出发点可能存在的“选择”；头节点表示了要达到

的“目标”。入度和出度总和最大的节点被称为网络的中心节点。中心节点表示了从“选择”到“目标”的关键节点。这样的贝叶斯网络也被称为因果网络或认知图(cognitive map)网络^[17-19]，如图3所示。

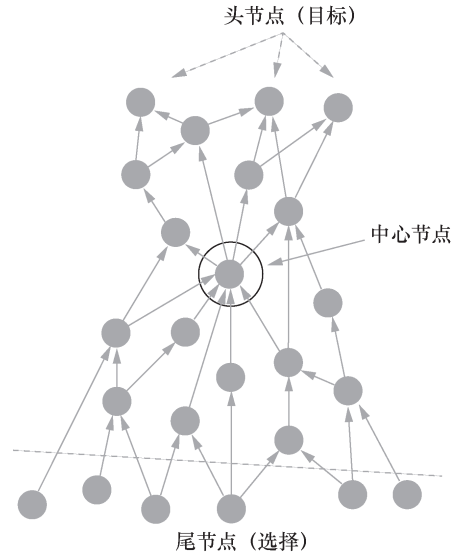


图3 认知图网络

4.2 人工神经网络法

人工神经网络法是用数学运算的方法给出神经元的输入和输出之间的关系，人工神经网络模型如图4所示。图4(a)给出了单个的神经元模型，其输出 y 与输入 $x_1, x_2, \dots, x_n$ 的关系如下：

$$y = f\left(\sum_i w_i x_i + b\right) \quad (2)$$

其中， w_i 为输入 x_i 的权重系数， b 为神经元的输入偏置， $f(\cdot)$ 为非线性函数。典型的非线性函数为Sigmoid函数：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

或符号函数：

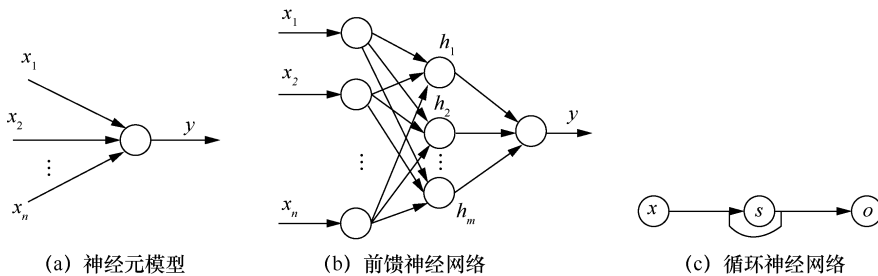


图4 人工神经网络模型

$$f(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (4)$$

由多个人工神经元组成的网络被称为人工神经网络。如果从输入到输出,神经元按照层次排列,层内神经元之间没有连接,层与层之间的神经元有单向连接,则这样的人工神经网络被称为前馈神经网络,如图4(b)所示。多于3层的人工神经网络被称为深度神经网络。如果神经网络内部具有反馈环路,则这样的网络被称为循环神经网络(recurrent neural network, RNN),如图4(c)所示。RNN具有记忆功能,适合用于序列处理^[18,20]。

人工神经网络具有较强的学习能力和推理能力。给定神经网络一定批量的输入和输出,可以训练神经网络各层的权重系数和偏置。这个过程一般采用梯度下降法,训练得到的参数符合最大似然估计。当有了权重和偏置参数之后,给定神经网络的输入就可计算人工神经网络的输出。人工神经网络的输出可以是连续输出,也可以是离散输出,这取决于输出层采用的非线性函数。连续输出可用于数据拟合,离散输出可用于分类^[18,20]。

4.3 脉冲神经网络法

脉冲神经网络法是一种与人工神经网络相似,但比人工神经网络更贴近生物物理的一种神经网络建模方法。脉冲神经网络与人工神经网络的差别如下。

① 神经元模型一般使用(偏)微分方程,具有时间参量,例如式(5)是一个典型的脉冲神经元模型。其中, V_m 为膜电压, C_m 为膜电压的跨导电容, R_m 为膜电阻, i 、 k 分别为兴奋性和抑制性神经细胞集合内的序号, I_i^e 、 I_k^i 分别为第 i 个兴奋性和第 k 个抑制性神经细胞的突触电流, w_i^e 、 w_k^i 为突触强度。

$$C_m \frac{dV_m(t)}{dt} = \sum_i w_i^e I_i^e + \sum_k w_k^i I_k^i - \frac{V_m(t)}{R_m} \quad (5)$$

② 神经元与神经元之间传递的是神经脉冲,为离散值,且神经脉冲没有幅度特性(即脉冲幅度为常值)。

③ 突触模型只能为正值。

④ 神经元模型被分为激励性神经元和抑制性神经元。

由于神经元与神经元之间传递的是神经脉冲,这里就存在一个神经脉冲如何表达信息的问题。目前学术界提出的神经脉冲对信息的编码方式有如

下几种:①频率编码(frequency coding),即使用神经脉冲的频率来表达信息;②时间编码(temporal coding),即使用神经脉冲出现的时间或脉冲与脉冲之间的相对时间间隔来表达信息;③群体编码(population coding),即使用一个神经元集群的平均放电率(即任意一个时刻放电的神经元数目与集群内神经元总数目的比)来表达信息。目前,对于大脑中到底采用何种方式来表达(或传递)信息,并没有明确结论。有证据表明,不同的大脑区域可能采用不同的编码方式。频率编码是目前被发现的大脑区域使用较多的一种方式。但频率编码的缺陷是它的反应速度较慢。一个神经元需要较长时间(例如1s)才能统计出它的输入神经元的放电频率。如果是一个深层网络,则需要更长时间才能从网络的输入端反应到输出端。这显然与大脑的行为是相悖的。例如,从大脑的视觉输入到行为输出经过了多层网络,但有些行为的反应非常快速。与频率编码和时间编码相比,群体编码的计算速度非常快。因此,很多人认为大脑内普遍采用的是群体编码,而群体内的每个神经元都采用频率编码方式^[13,21-23]。

4.4 神经动力学法

神经动力学法是一种比脉冲神经网络法更通用的建模方法。正如理论物理四大力学都使用(偏)微分方程建模一样,神经动力学也使用(偏)微分方程,对神经元、树突、轴突、神经纤维等进行建模。神经动力学建模观察的往往是网络的整体特征,该整体特征可能代表了某种信息。例如大量神经元的同步放电(兴奋)现象可能代表了多个特征的绑定(binding)^[24]。

4.5 各种方法的比较

以上各种方法都可用于大脑的建模,但各种方法都有各自的特点。贝叶斯网络法属于通用建模方法,在其他领域也有广泛应用,例如气候变化建模。人工神经网络法和脉冲神经网络法都是专属于神经网络的建模方法。人工神经网络比较抽象,但具有更好的数学特征和学习能力。脉冲神经网络更接近于人脑的生物物理特征。动力学方法也属于通用建模方法。神经动力学方法是动力学方法在神经科学里的具体应用。神经科学的任何部分,包括神经细胞、树突、轴突、神经纤维等都可以用动力学方法建模。

这里有一个很有趣的问题:贝叶斯网络法使用概率/随机的方法对大脑的行为进行建模,给定相同

的内部状态和外部输入，仍然可能产生不同的外部输出；神经动力学法使用微分方程对大脑的行为进行建模，给定相同的初始条件和输入，一定产生相同的输出（即决定论）。那么大脑的行为到底是确定的，还是随机的，目前科学界并没有定论。但是，由于脑科学中仍然有很多问题没有解决，使用完全确定的模型是行不通的。因此，在采用神经动力学法建模时，一般会增加噪声项，以补充对未知因素的描述。

5 大脑的框图模型

框图是一种简单的描述模型的方法。本节从计算机的框图模型开始，类比给出大脑模型。

5.1 从计算机模型到大脑模型

计算机的模型如图 5 (a) 所示，分为输入、存储/计算和输出三大部分。类似地，可以给出大脑模型，如图 5 (b) 所示，分为感知、记忆/决策和行为。感知部分包括各种感觉器官，例如视觉、听觉、嗅觉、触觉和味觉等。记忆/决策部分主要指大脑的前额皮层、海马和杏仁核。前额皮层实现人脑高级的思考、规划和语言等；海马与人的记忆密切相关；杏仁核与人的情感和动机密切相关，它可以确定大脑目标状态的选取，并且调节大脑的学习速率。行为部分主要与大脑的运动皮层相关。运动皮层位于大脑顶叶皮层，负责驱动运动神经，与肌肉相互作用。

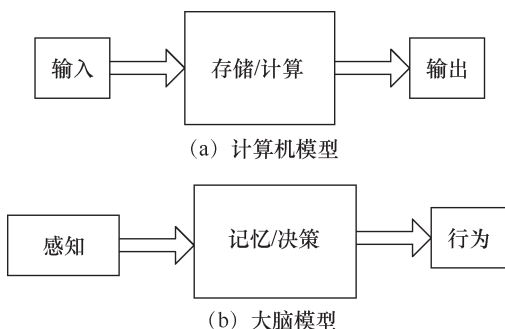


图5 计算机模型与大脑模型

记忆/决策部分是大脑有意识活动的主体，它接收来自感知部分的输入，从多个可能的动作中做出选择，并将选择的动作输出给行为部分进行执行。从心理学的角度，基于一个复杂策略的决策是心理活动的最高形式。通常用强化学习对决策部分进行建模。这基于两点理由：第一，强化学习也是在每一个状态，从多个动作中选择一个动作；第二，

强化学习优化的一个（复杂）策略的总回报，而不单纯是当前动作的回报。

对于感知部分，由于感知部分尚未进入大脑意识，无法给该部分的输出打标签或提供评价信息，因此该部分的学习只能为非监督学习。行为部分较简单。由于决策部分给出了目标信号，因此行为部分的学习是一个监督学习的过程。一般，感知部分对应数据处理上的降维变换，行为部分对应数据处理上的插值变换。

5.2 从感知到行为的直接通路

但是，人类行为学研究表明，并不是任何行为都需要思考和记忆。有些行为具有从感知到行为的直接通路，如图 6 所示。例如，一个飞石突然从你的眼前飞过，你自然地避让，这是不需要思考和记忆来决策的。在飞石进入你的意识层面之前，你已经行动起来避让飞石了。有科学家认为，这是人类为了生存和安全进化出来的基本能力，由基因决定，是与生俱来的，不需要学习和思考，也无须记忆（此时记忆是滞后于行为的）。图 6 (a) 展示了这种模型，“基因决定的快速行为”即这种基因决定的、从感知到行为的直接通路。

还有一类行为也具有从感知到行为的直接通路，但却是需要学习的。例如打乒乓球，经过反复练习，看到乒乓球飞来，人自然知道怎样去挥动球拍接球，人的行为往往也是在思考决策之前就做出的。因此，这里是从视觉输入到行为输出的直接通路。与前述行为的差别是，这一类行为需要经过反复练习得到。图 6 (b) 描述出了这二者的差别，“习得的快速行为”即这种学习获得的、从感知到行为的直接通路。

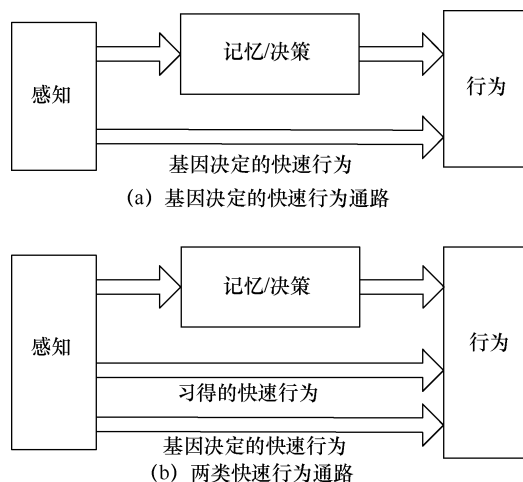


图6 从感知到行为的直接通路

5.3 分布式自适应控制模型

分布式自适应控制模型是目前科学界提出的比较详细的大脑框图模型^[25-26]。分布式自适应控制模型如图7所示。在图7中,最左侧从下到上的一列是从感知输入到进入大脑意识的通路;中间从下到上的一列是从身体的生理需求输入到进入大脑意识的通路;最右侧从上到下的一列是行为选择通路。行为可以由大脑意识层面产生,也可以由感知及生理需求直接产生。感知及生理需求产生的行为又被分为反射式行为和自适应式行为两种。反射式行为即基因决定的、人自发的行为;自适应式行为即学习获得的、从感知到动作输出的快速行为。

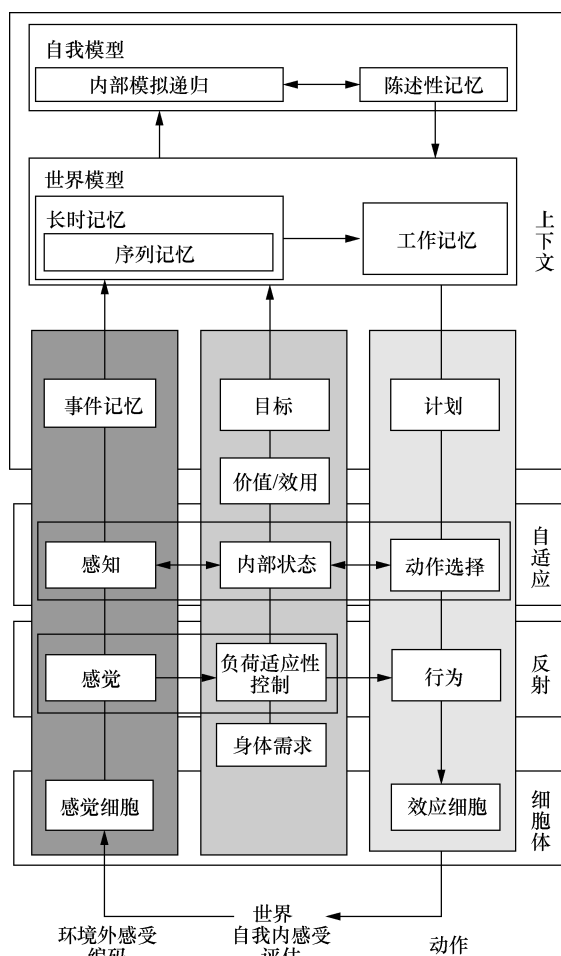


图7 分布式自适应控制模型

在意识层面,大脑主要负责记忆和决策。大脑会建立两个模型:“世界模型”和“自我模型”。“世界模型”主要反映外部的客观世界,包括外部的一切客观事物、别人以及外部世界各个人、物之间的关系等。因此,“世界模型”自然也包括大脑对外

部社会(即人与人之间关系)的认知。“自我模型”主要描述大脑对人自己决策和能力的评估。例如,“我确信”是对自己决策信心的评估,“我擅长数学”“我擅长运动”是对自己能力的评估。对决策信心的评估会影响后续决策。例如,信心高,则决策者后续会增加“利用”先前经验的成分;信心低,则决策者后续会在自己的决策中增加“搜索”的成分^[27]。面对一个复杂任务,决策者会将复杂任务分解为多个子任务。对自己能力的评估会影响子任务的划分和选取。例如,假如需要在20 min内到达某地,是“跑步”去还是“骑自行车”去?在行动之前,需要对自己执行两种策略(“跑步”和“骑自行车”)的能力有一个评估。这种自我能力的评估结果会影响决策(即选择“跑步”策略还是“骑自行车”策略)。由此可见,“自我模型”主要是对自己“认知”活动的更深层“认知”,因此被称为“元认知”^[28-29]。

意识层面基于“世界模型”和“自我模型”,制定出一套行为计划,该计划指引行为,对外部世界施加动作,改变外部世界的状态,使得外部世界逼近自己的感知和生理需求。

需要说明的是,图7右侧从意识层面出发的行为有时会与从感知层面出发的快速行为出现竞争/抑制现象。例如,在冬天将手放进一盆凉水里,从感知出发的行为会指导你将手拿出;但是,假如在一个游戏规则里,你放进去5 min将得到一定奖励,你的意识层面可能会抑制将手拿出的行为。这种竞争/抑制行为可以由赢者通吃(winner-take-all, WTA)网络来实现^[30]。

还需说明的是,图7中从感知到行为的直接通路在物理实现上不一定在大脑内部,而可能位于神经系统的其他部分,这里不加强调和区分。例如,行为部分往往对应数据处理上的插值变换,这在很多情况下是由位于脊椎内的脊髓神经来实现的^[11-12]。脊髓内的中枢模式发生器(central pattern generator, CPG)可以产生规律的波形,协调多处肌肉的联动,产生协调的动作行为^[31-33]。例如,走路涉及手臂、大腿、小腿等肌肉的联动,这就是由CPG来协调控制的。而大脑的意识部分只需要对CPG下达一个简单的命令,无须关注运动控制的细节。此外,有些CPG参数还可受到意识部分的调节,例如可以由调节CPG产生波形的频率来控制走路快慢。

6 大脑的数学模型

6.1 自由能量原理及其内涵

自由能量原理来源于热力学第二定律。热力学第二定律告诉人们，一个系统的内部和外部总是力求达到平衡，即减小系统内部和外部之间的差别。大脑是一个信息处理系统，因此，大脑总是力求减小大脑内部与外部之间的互信息，即减小吃惊程度^[34-35]。

大脑外部的信息 $-\log p(x)$ 是指从人的感觉器官输入大脑的信息；大脑内部的信息 $-\log p(y)$ 是指大脑内部对外部输入的预期。大脑总是力图减小二者之间的互信息 $I(x, y)$ ：

$$I(x, y) = -\log \frac{p(x)p(y)}{p(x, y)} \quad (6)$$

可以由机器学习中常见的自动编码器或波尔兹曼机实现式 (6)。自动编码器和波尔兹曼机的原理如图 8 所示。

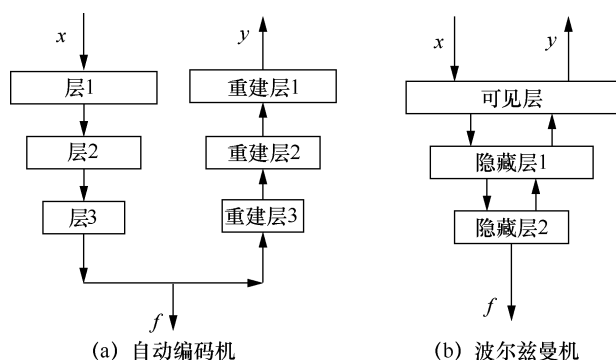


图8 自动编码器和波尔兹曼机的原理

在图8(a)中，外部世界的输入 x 通过几层神经网络提取输入的特征 f ，然后该特征往回反馈重建，输出对外部世界的预期 y 。自动编码器总是力图减小 x 和 y 之间的误差。在图8(b)中，每一层分为可见节点和隐藏层节点。可见节点的输出 y 总是力图重建输入 x 的概率分布 $p(x)$ ，即使得 $p(y)$ 与 $p(x)$ 之间的误差尽量小，尽量最小化式 (6)。

自动编码器和波尔兹曼机^[36-38]都属于非监督学习。人脑感觉器官的功能可以被建模为采用这种形式的神经网络，来提取外部世界的特征。这些特征将进入人脑的意识层面，一方面进入记忆模块保存起来，另一方面即时地参与决策。

此外，自由能量原理还可以解释人类的行为和动机。如图7所示，外部世界对大脑的输入不但包括感觉细胞的感知输入，也包括身体的生理需求输

入。这些生理需求将联合大脑“自我模型”产生的心理需求，一起产生人的深层次的“动机”。深层次的“动机”将结合感知输入生成大脑短期行为的“目标”。“目标”指导行为，行为施加在外部世界上，改变外部世界的状态，使得外部世界尽量满足大脑的生理需求和心理需求，减小二者之间的互信息^[39]。

然而，外部世界遵循它自己的规律。对外部世界施加的动作并不能使得外部世界立即达到大脑期望的理想状态，二者之间经常会有一定的误差。这个误差将反过来作用在大脑上，从而引起大脑的缓慢变化。这就是大脑的可塑性。

6.2 大脑神经网络的基本模块：WTA 网络

WTA 网络的工作原理如图9所示。在图9中，输入（例如，通过自动编码器或波尔兹曼机提取出来的特征）可以用向量来表示，图9中同时展示了原有的两个向量 W_1 和 W_2 。当输入向量 **Input** 与向量 W_2 较接近时（此时需要竞争比较），输入将被判决为 W_2 （即 W_2 为赢者）。同时，原有的 W_2 将与输入相互作用，如式 (7) 所示形成新的 W_2' 。其中 λ 为更新率，而向量 W_1 保持不变。

$$W_2' = (1 - \lambda)W_2 + \lambda \cdot \text{Input} \quad (7)$$

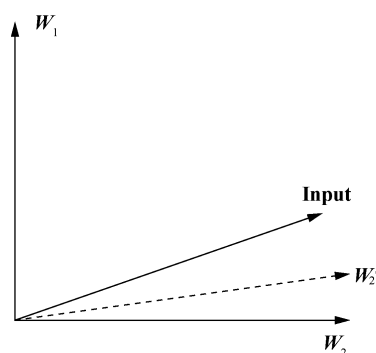


图9 WTA网络的工作原理

如果一个WTA网络有 K 个向量，随机初始化各向量，当输入数据具有一定的聚类特征（如图10所示）时，随着输入的进行，网络的各个向量将自动聚类，逐渐形成一个向量表示输入的一个类别的情形。因此，WTA网络具有自动聚类的功能。它属于一种竞争网络，实现的是一种非监督的学习方法^[13,30]。

从神经网络实现形式来看，WTA网络一般为RNN。例如著名的霍普菲尔德（Hopfield）网络^[20]就是典型的WTA网络。WTA网络是大脑记忆的一个较好的模型。

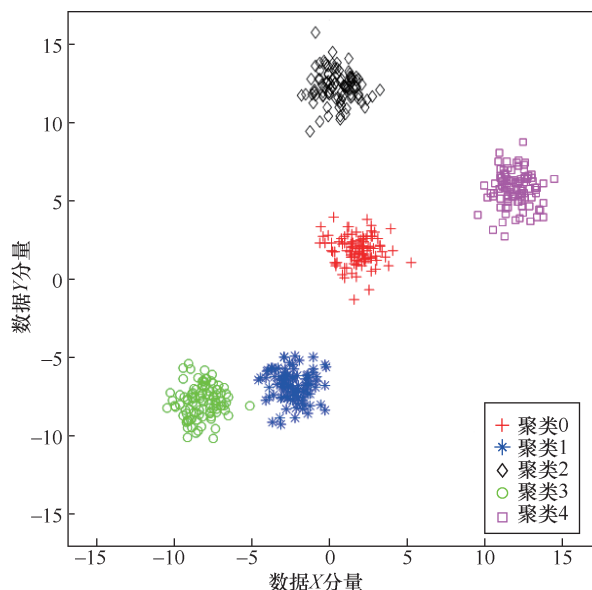


图10 具有聚类特征的输入数据

WTA 网络的聚类功能具有“越经历、越强化”的特征，这符合人类大脑的行为特征。有研究认为，WTA 网络可能是构成人脑的基本模块^[12,13,40-42]。人类大脑皮层从上到下分为6层，大脑皮层内神经元的连接方式主要为上下层之间的纵向连接，但是在第3层或第4层，具有层内神经元的侧向连接。这被认为是大脑皮层内神经元构成 RNN 的标志。这与 WTA 网络的实现形式是一致的。

6.3 强化学习和大脑

强化学习是一个马尔可夫决策过程。马尔可夫决策过程 M 可以用一个五元组 $\langle \chi, A, R, P, P_0 \rangle$ 来表示。其中 χ 为状态集， A 为动作集， $R(x, a)$ 为在状态 x 下采取动作 a 时的回报， $P(\cdot | x, a)$ 为状态转移概率， $P_0(\cdot)$ 为各状态的初始分布概率。记一个静态策略 $\mu(\cdot | x_0)$ 为从某一状态 x_0 开始的一系列动作的集合。对于一个给定的静态无穷策略 μ ，定义总回报 D^μ 为：

$$D^\mu(x) = \sum_{t=0}^{\infty} \gamma^t R(x_t, a_t) | x_0 = x, \mu \quad (8)$$

强化学习解决的问题是求解从状态 x_0 出发且取得总回报 D^μ 最大的策略 μ 。

设定一个时间限制，设最大动作转移步骤（即时间）为 T ，设定一个目标状态 x_d ，在时间 T 内到达目标状态 x_d ，则总回报如式（8）所示，否则总回报为 0。则这样的马尔可夫决策过程可求解一个从初始状态 x_0 到目标状态 x_d 的最优路径。

6.3.1 强化学习的求解过程

目前常见的强化学习求解过程采用迭代收敛方法，包括值迭代方法和策略迭代方法。基于值迭代的强化学习求解过程如图 11 所示。在图 11 中，在每个状态 s 上，附着一个值参数 $V(s)$ 。 $V(s)$ 的物理意义为从状态 s 开始，到达目的节点 D （即 x_d ）的期望总回报。任意状态 s 的 $V(s)$ 的初始值都可随机初始化。不断采用贝尔曼公式对 $V(s)$ 值进行迭代：

$$V(s) = E_{s'} \{ R(s, s') + \gamma V(s') \} \quad (9)$$

其中， $E_{s'}(\cdot)$ 为对状态 s' 求期望， $R(s, s')$ 为从状态 s 转移到状态 s' 时的回报， γ 为衰减因子。随着迭代的进行，各个状态的 $V(s)$ 值逐渐收敛，最后趋于稳定。最后，从源节点 S 到目的节点 D 的最优（回报最大）策略为：

$$V^*(s, a^*) = \max_a \{ R(s, a, s') + \gamma V(s') \} \quad (10)$$

$$\mu(s) = a^*$$

其中， $R(s, a, s')$ 为在状态 s 下采取动作 a ，到达状态 s' 的回报。

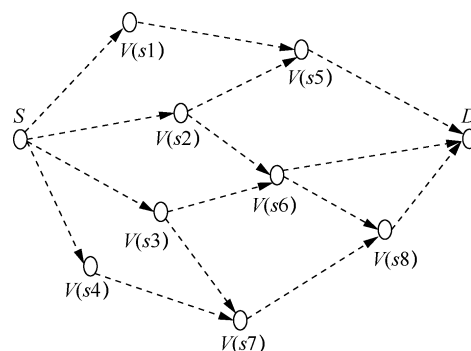


图11 基于值迭代的强化学习求解过程

将式（9）进行变形，得到时间差分信号：

$$T(s, s') = V(s) - V(s') \quad (11)$$

和时间差分误差信号：

$$\delta(s, s') = V(s) - V(s') - R(s, s') \quad (12)$$

基于策略迭代的强化学习求解过程如图 12 所示。在每个状态 s 上附着一个参数 $p(s, a)$ 。 $p(s, a)$ 的物理意义为在状态 s 上选择动作 a 的概率。任意 $p(s, a)$ 的初始值都可随机初始化。学习过程如下：从源状态开始，在每一个状态按照各个动作的概率分布随机选择一个动作，如此形成一个策略 μ 。如果在 T 步以内（含）到达了目标状态，则可以积累每一个状态转移的回报，得到总回报 $D(\mu)$ ；否则， $D(\mu) = -\inf$ ，其中 $\inf$ 为无穷大。

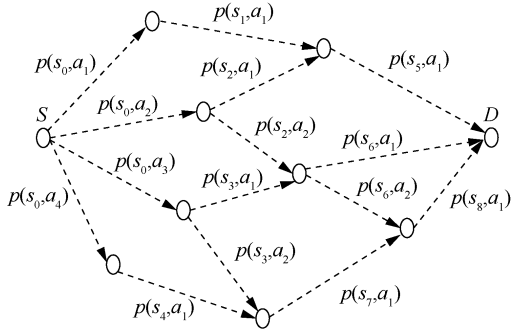


图 12 基于策略迭代的强化学习求解过程

设置一个最优回报 D^* ，并初始化 $D^* = -\inf$ 。对于每一个策略 μ ，如果 $D(\mu) \geq D^* \wedge D(\mu) \neq -\inf$ ，则策略 μ 加强，并更新 D^* ：

$$\begin{aligned} p(s, a) &= p(s, a) + \Delta, \forall (s, a) \in \mu \\ D^* &= D(\mu) \end{aligned} \quad (13)$$

其中， Δ 为一个小的正值。否则，策略 μ 减弱，如式 (14) 所示，无须更新 D^* ：

$$p(s, a) = p(s, a) - \Delta, \forall (s, a) \in \mu \quad (14)$$

然后，重新归一化各状态的动作概率：

$$p(s, a_k) = \frac{p(s, a_k)}{\sum_i p(s, a_i)} \quad (15)$$

经过充分学习后，最优策略为在每一个状态选择动作概率最大的动作，即：

$$\begin{aligned} p(s, a^*) &= \max_a p(s, a) \\ \mu(s) &= a^* \end{aligned} \quad (16)$$

6.3.2 强化学习的神经科学、心理学、行为学依据

(1) 基于值迭代的强化学习的神经科学、心理学依据

首先，在值迭代方法中， $V(s)$ 的物理意义为从状态 s 到目标状态的期望总回报。“期望”本身就是一个心理学概念。按照式 (10) 选择最优策略，即总是选择期望回报最大的策略，这也是符合人的心理行为的。

从神经科学的角度来看，现在已有稳定的神经科学实验表明，中脑腹侧被盖区 (ventral tegmental area, VTA) 多巴胺神经元的位相活动编码了值迭代的时间差分预测误差，即式 (12)。此外，灵长类动物外侧缰核部位的神经元活动在相反方向上反映了时间差分预测误差信号。同时，在纹状体、杏仁核、眶额皮层 (orbitofrontal cortex) 和前扣带皮层都可以观测到与期望总回报 $V(s)$ 相关的神经元活

动。这些区域可能编码了值函数中的“状态值”^[41,43]，这些“状态值”从感知和记忆中得来。

这里仍然有一个问题：多巴胺作为一种神经递质，存在于多巴胺神经元的突触当中，中脑腹侧被盖区有很多个多巴胺神经元，也有很多个突触，于是多巴胺神经元的位相活动就有很多份表达，这可以用一个向量来表示，然而选取目标后，根据式 (12)，大脑中只有一个值迭代的时间差分误差信号，是一个标量，那么这个向量和这个标量究竟是如何对应的呢？参考文献[44]提出，这个向量以一种概率分布的方式对这个标量进行了表达，这是前述群体编码在“期望回报误差信号”编码方式上的具体实现方式。

(2) 基于策略迭代的强化学习的神经科学、行为学依据

当产生“好”的结果时，策略迭代的策略加强；当产生“不好”的结果时，策略迭代的策略减弱。这是符合人的行为的。经过充分学习后，概率越大的动作重复越频繁，重复越频繁的动作产生的结果越好。此时，人只是根据当前状态各动作的概率（频繁程度）来选择动作，而没有一个“目标”的概念（即选择动作时不考虑目的节点 D ），也就是说，这个动作完全是人的“习惯性”动作^[45-46]，而非“目标驱动”。策略迭代学习到一个好的策略，相当于人养成了一个好的“习惯”。人的行为既由“目标驱动”，也由“习惯驱动”。大脑会对这两种驱动信号进行合并或选择。当“习惯驱动”占主导时，“目标驱动”会减弱，这种现象被称为“去值化”^[47]，即去掉“目标驱动”中的“期望值”。

(3) 无模型的强化学习和基于模型的强化学习

在基于策略迭代的强化学习过程中，只是根据当前状态 s 的所有可能动作进行每一步的动作选择，而无须知道选择该动作后将跳转到哪一个状态 s' 。也就是说，大脑在选择动作时，无须知道图 12 所示的路径转移结构，即无须知道外部世界的“世界模型”。这种方法被称为无模型的强化学习。

在图 11 所示的基于值迭代的强化学习求解过程中，每一步的动作选择（如式 (10) 所示）需要利用到动作 a 将导致的下一个状态 s' 。因此，基于值迭代的强化学习过程是需要知道图 11 所示的路径转移结构的。这种方法被称为基于模型的强化学习。此外，在选择动作时，需要利用 $V(s')$ ，而 $V(s')$ 是与目的节点 D 相关的。这种学习过程也被称为

“目标驱动”的强化学习^[48-52]。

(4) “世界模型”的建立

最初大脑对外部世界一无所知，只能采用无模型的强化学习。在学习过程中，人们一方面会“真实经历”从状态 s 经过动作 a 转移到状态 s' 的过程，由此可以训练一个（神经网络）模型：

$$s' = f(s, a) \quad (17)$$

另一方面，随着学习的进行，在每一个状态 s 转移到 s' 时，都有一个回报值 $R(s, s')$ 与之对应，同样可以训练一个回报值的（神经网络）模型：

$$R(s, s') = g(s, s') \quad (18)$$

式 (17) 和式 (18) 中， $f(\cdot)$ 和 $g(\cdot)$ 表示函数，它们反映了外部世界的转移结构以及路径转移所带来的回报值。式 (17) 和式 (18) 被统称为“世界模型”^[53-54]。通过“真实经历”，学习式 (17) 和式 (18) 的模型，相当于在人脑中建立了一个“世界模型”。这个过程也被称为“建图”。

(5) 前瞻性思考和回顾性思考

① 前瞻性思考

（神经网络）模型具有外推能力。于是，对于没有“真实经历”的状态 s ，通过式 (17) 和式 (18) 可以预测出 s' 的值。从 s 到 s' 的转移，没有真实经历，只在头脑中进行，这相当于人的思考过程。反复利用式 (17) 和式 (18) 进行“思考”，可以得到一个通过“思考”得到的路径转移轨迹，这个轨迹被称为“虚拟经历”。

采用“虚拟经历”反复进行搜索，即采用“虚拟经历”进行无模型的强化学习，这可以用蒙特卡洛树搜索（Monte Carlo tree search, MCTS）来模拟^[55-57]。而 MCTS 得到一个好策略，就相当于“思考”得到一个好结果。

MCTS 的特点是搜索速度较慢。但对于未曾经历的现象，人脑做决策思考的速度慢也是很自然的事情。从这个角度来说，MCTS 能较好地模拟大脑的前瞻性思考。

② 无模型的回顾性思考

对于有“真实经历”的事件，除了建立式 (17) 和式 (18) 的“世界模型”，还可以直接利用式 (13)、式 (14) 获得状态 s 下动作的概率值，做出动作选择。这个过程非常简洁、快速。对于准确度要求不高的场合，可以用这种机制来模拟大脑快速做出决策的过程。

③ 基于模型的回顾性思考

对于有“真实经历”的事件，大脑还可以根据

式 (9)，将回报值 $V(s)$ 逐渐由后向前反馈（即由 $V(s')$ 传递到 $V(s)$ ），这将涉及一个反向推理的过程^[58-59]。对于不同步数的时间差分，反向推理的复杂度不同。甚至在极端情况下，不记忆中间节点的 $V(s')$ ，直接由终点的 $V(D)$ 倒推至当前状态的 $V(s)$ ，这个过程将涉及相当大的计算量。但与前瞻性思考和无模型的回顾性思考相比，这个过程计算量居中，因此大脑的思考决策速度也居中。

(6) 思考：快与慢

① 两套系统

以上 3 种思考决策方式中，无模型的回顾性思考最快，前瞻性思考最慢，基于模型的回顾性思考居中。简单起见，称无模型的回顾性思考为快速决策系统，称前瞻性思考和基于模型的回顾性思考为慢速决策系统。于是可以简单地说，无模型的强化学习是一个快速决策系统，基于模型的强化学习是一个慢速决策系统（请注意，前瞻性思考也是基于模型的）。大脑的思考决策正是这两套系统联合达成的。

诺贝尔奖获得者丹尼尔·卡尼曼（Daniel Kahneman）在其著作《思考：快与慢（thinking: fast and slow）》中，将大脑的思考归类为两套系统：系统 1 和系统 2^[60]。系统 1 快速、直接、勤奋，时时刻刻在工作，但是容易受到情感、情绪、社会因素等外部因素的影响，容易犯错。系统 2 理性、深入、准确度高，但是系统 2 懒惰、慢速、对脑力（能量）的消耗大。Daniel Kahneman 的系统 1 和系统 2 是在心理学层面对大脑思考过程工作机理的定性描述。在数学机理层面，无模型的强化学习和基于模型的强化学习可以分别建模为系统 1 和系统 2 的具体实现方式。

② 两套系统的交互作用

需要强调的是，无模型的强化学习和基于模型的强化学习是相互作用的。一方面，式 (17) 和式 (18) 表明，在无模型的强化学习过程中，大脑会建立“世界模型”，从而为基于模型的强化学习奠定基础。另一方面，在后续无模型的强化学习过程中，大脑可以根据已经习得的“世界模型”对无模型的强化学习（随机搜索）进行偏置（即作为先验知识），从而减小无模型的强化学习的搜索空间，提高搜索效率^[58,61-62]。因此，无模型的强化学习和基于模型的强化学习是交织在一起的。正是这二者的相互作用构成了大脑智慧的源泉：无模型的随机搜索构成了大脑的灵感；基于模型的强化学习提高了准确

度,从而使得很多的“思考”具有可行性^[58-59,63-64]。随着二者的交织进行,一方面,“世界模型”的内涵越来越丰富(即式(17)和式(18)的准确度越来越高);另一方面,基于模型的先验知识越来越准确,无模型的搜索空间不断减小,这使得很多原本无法达成的事情得以实现。

(7) 巴甫洛夫学习

巴甫洛夫学习也称经典学习,是俄国生理学家巴甫洛夫发现的一种生理学现象,其实验表现如下。每次给狗喂食物之前,都响一声铃铛,然后才给狗食物。而狗一见到食物,就会流口水,做好进食的准备。经过长期训练后发现,响一声铃铛,即便不给狗食物,狗也会流口水,做好进食的准备^[13]。

巴甫洛夫学习是与操作性学习相对的。操作性学习指如下现象。在一个实验箱内,小鼠可以按动两个操作杆。小鼠按动左侧操作杆,则会得到食物;按动右侧操作杆,则会得到水。经过训练,小鼠就会知道:饿了,就去按动左侧操作杆;渴了,就去按动右侧操作杆。

巴甫洛夫学习与操作性学习的区别是:在巴甫洛夫学习中,狗学会的是外部世界的变化,从而对外部世界的变化走向做出了预测;在操作性学习中,小鼠学会了如何对外部世界施加激励,驱动外部世界的变化,以便满足自身的期望(需求)。

巴甫洛夫学习告诉人们,人脑不但会对外部世界建立一个“世界模型”,还会学习“世界模型”的导数,由此对外部世界的变化做出预测^[35,53]。

巴甫洛夫学习的建模同样可以采用式(9)和式(10)进行。不过此时的动作是外部世界的变化导致的,而不是生物体所实施的。

(8) 强化学习小结

从上文分析可见,强化学习主要是对大脑的意识(决策/思考)部分进行建模,它较好地将大脑的各种不同决策/思考方式统一在一个框架之下。图13给出了基于强化学习的大脑决策模型。在图13中,巴甫洛夫学习、无模型的强化学习和基于模型的强化学习是并行执行的^[53]。它们的输出转化为统一的值估计,并都连接到运动皮层,最后由运动皮层选择或合成这些信息,从而生成最终的动作。现在普遍认为,意识部分并非只有这3种并行模块^[46]。但其他并行模块是什么以及它们如何工作,仍然有待进一步的研究。对于运动皮层中控制同一块肌肉的运动神经元集群,多个并行通道的值估计可以以式(19)对

其合并关系进行建模,即对多个通道的值估计进行求和^[65]:

$$Q(s) = \sum_i Q_i(s) \quad (19)$$

强化学习模型给人们提供了一个认识大脑功能、认识实验数据的方法。可以利用第2节中的实验手段对强化学习模型进行验证。实验表明,大脑的决策/思考部分功能主要集中在大脑的前额皮层;“世界模型”主要集中在大脑的前额皮层和海马区;而强化学习的目标状态由大脑的杏仁核来编码^[66-67]。

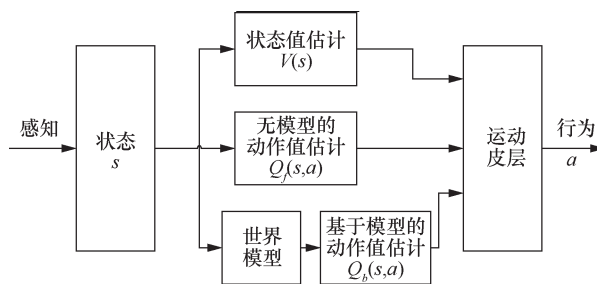


图13 基于强化学习的大脑决策模型

6.4 情感建模

情感是人对客观事物是否满足自己的需要而产生的态度体验^[68]。它经常取决于人们对客观事物的认知^[69]。对于相同的客观事物,不同的人可能会产生不同的认知,因而产生不同的情感。因此,情感和大脑的意识是紧密联系在一起的。实验表明,情感主要由大脑内的杏仁核产生和处理。

情感是大脑内的又一个并行处理通路。它包括情感识别、情感决策和情感表达,分别对应大脑学习通路的感知、决策和行为部分^[70]。并且,情感通路的情感识别、情感决策和情感表达与学习通路的感知、决策和行为是紧密耦合在一起的。情感识别和情感表达因其应用性较强,获得了较多的关注,这里不再赘述。本节主要关注情感决策。

情感决策是指情感对大脑行为决策的影响。情感决策和大脑学习通路中的学习决策一起决定了大脑的行为决策,从而指导最终的行为输出。情感决策与学习决策的关系如图14所示。在图14中,感知输入经丘脑同时进入各种感觉皮层和杏仁核。各感觉皮层的信息联合眶额皮层的回报信号,形成决策信息。同时,决策信息受到情感(杏仁核)的影响,并反过来影响情感(杏仁核)。最后,决策和杏仁核内的情感联合支配了运动皮层的行为。图14的模型被称为基于大脑情感学习的智能控制器

$$LR = k \frac{1}{\text{conf}(s)} = k\sigma(s) \quad (24)$$

其中, 抑郁症患者表现为对正面回报学习时, $k < 1$; 焦虑症患者表现为对负面回报学习时, $k > 1$ 。

6.4.3 情感对决策的其他影响方式

上述情感对决策的影响模型都比较简单, 但实际上, 情感是一个很复杂的模块, 它影响着大脑思考和决策的方方面面, 例如^[75]:

- ① 情感影响着人们思考的内容;
- ② 情感影响着人们思考的深度;
- ③ 情感影响着人们目标的选择;
- ④ 情感影响着人们的社交决策。

总之, 情感在决策中占有重要地位, 但目前对情感和决策的研究还很不成熟, 其模型也尚不清晰。

6.5 自然语言处理

这里的自然语言处理模块是指 PFC 内的布罗卡区 (Broca's area), 而不是感觉系统中的听觉皮层 (auditory cortex)。语言处理是人类大脑特有的一个重要特征^[11-12]。它位于意识层面, 是大脑内又一个并行处理模块, 且其独立性相对较强。自然语言处理模型由于其应用性较强, 近几年得到了极大的发展, 已发展成为一门独立的学科。这里只简要概述其模型架构, 并做出几点评价。

6.5.1 模型架构

目前流行的端到端自然语言处理模型^[76-77]如图 16 所示。图 16 中输入数据 $(x_1, x_2, \dots, x_n)$ 为一个向量序列, 它经过编码器之后输出隐变量序列 $(h_1, h_2, \dots, h_k)$, 该隐变量序列经注意力 $\alpha_{i,j}$ 加权后形成上下文向量 C_i 序列, C_i 序列经解码器解码之后输出目标数据序列 $(y_1, y_2, \dots, y_m)$ 。图 16 中的编码器、解码器分别为一个神经网络, 常用的网络形式有 RNN、CNN 和自注意力 (self-attention, SA) 模型等^[76-77]。

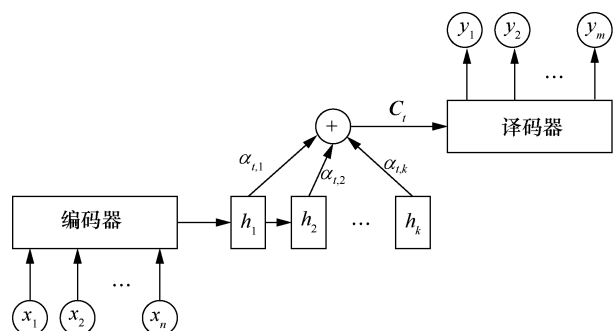


图 16 端到端自然语言处理模型

图 16 的模型可广泛应用于语音识别、机器翻译、摘要生成、机器问答等场景。所谓端到端, 是指在训练时, 直接在模型的输入端指定输入源数据, 在输出端指定输出目标数据, 进而进行训练。即使输入数据和输出数据属于不同的数据类型 (例如输入语音信号, 输出文本信息), 也无须添加特定的特征提取层。

6.5.2 模型的进一步扩展

受图 16 端到端架构的启示, 参考文献[78]给出了一个从语言序列到动作序列的端到端架构。并且, 该架构具有双向功能, 即它还能从动作序列反过来生成语言序列。从语言理解到动作序列的输出架构如图 17 所示。该架构在语言侧和动作侧分别采用多种时间尺度的长短期记忆 (multiple timescale long short-term memory, MT-LSTM) 抽取特征, 通过关联层 (binding layer 或 meaning layer) 将语言和动作序列关联起来^[79], 实现二者的端到端翻译。但目前该架构只在较小的语言和动作序列集上取得了成功。

6.5.3 模型评价

(1) 图 17 的端到端且独立于大脑其他功能的架构, 很好地诠释了大脑的并行处理架构。

(2) 图 16 中的注意力模型有一定的神经科学和心理学基础。在神经科学中, 注意力被解释为偏置竞争 (bias competition), 即从神经信号通路的一级到下一级, 神经资源 (例如神经细胞) 减少, 因此在信息的表示上产生竞争, 有些信息得到加强, 有些信息得到减弱 (甚至消失), 自下而上 (从感知到意识) 的偏置竞争形成显著性 (saliency)^[15], 自上而下 (从意识到感知) 的偏置竞争形成注意 (attention)^[80]。

(3) 图 17 的模型没有很好地与大脑的记忆、想象、思考等功能结合起来, 限制了模型的深度思考并作答能力, 这也是目前自然语言处理应用研究上面临的一个主要障碍。

(4) 图 17 的模型缺乏更深层次的生物物理依据。

(5) 图 17 的模型给了人们极大的鼓舞。这使得人们联想, 按照图 17 的模式, 不断增加模型规模和复杂度 (例如参考文献[72]中将视觉和语言理解融合在一起作为多模态感知侧), 是不是就可以构建一个虚拟大脑? 目前该模型在大数据集的泛化 (generalization) 上仍存在一定的困难。但即便是成功泛化, 仍然需要进一步探索生理层面的模型依

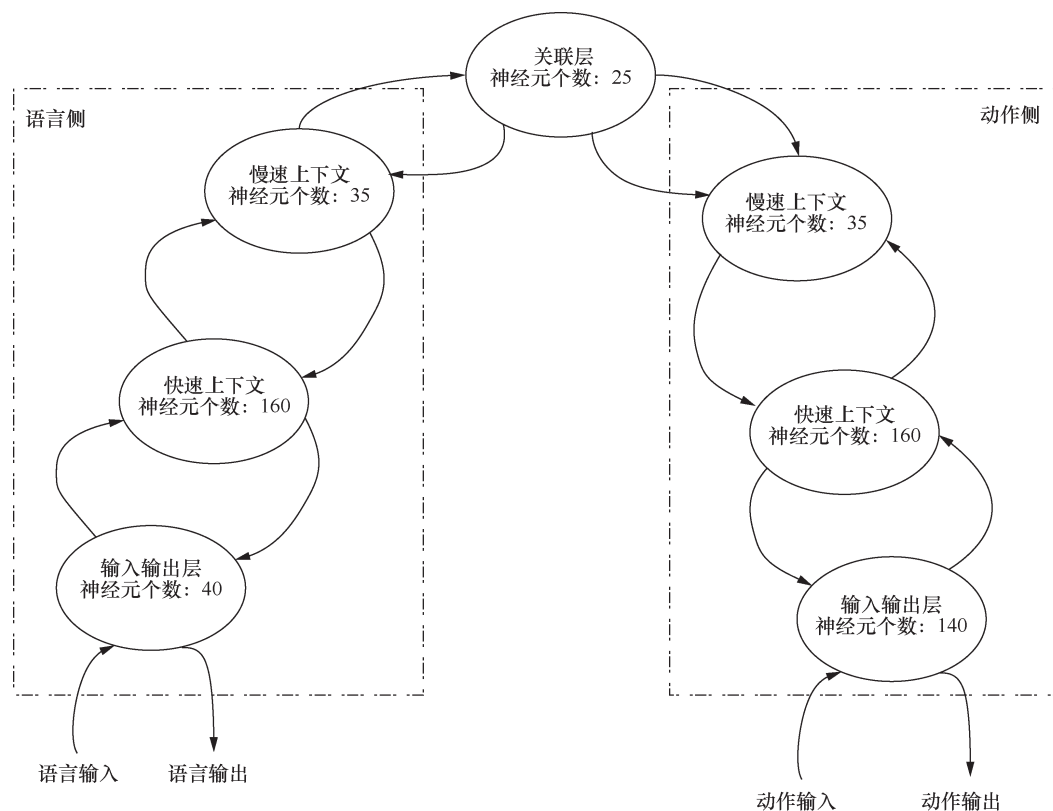


图 17 从语言理解到动作序列的输出的架构

据，以利于大脑疾病的诊断和治疗。

6.6 对“世界模型”的进一步讨论

从上文分析可以看到，“世界模型”在大脑建模中扮演着核心地位，因此这里对它做进一步的讨论。首先，“世界模型”是有心理学基础的^[81-82]。在心理学上，“世界模型”也被称为“认知图”。例如，一个家庭主妇在她熟悉的厨房里，不假思索就能知道冰箱在她身后，就是因为她脑子里早已建立好了这个“世界模型”。此外，这个“世界模型”的坐标是不依赖于她身体的朝向、身体坐标的。例如，现在冰箱在她身后，她闭上眼睛，转过身去，她能清楚地知道，冰箱在她前面。这说明，她脑中“世界模型”的坐标基准是独立于她的身体朝向的。有研究表明，该坐标基准的表达可能位于大脑顶叶皮层（superior parietal cortex, SPC）^[82]。

在神经科学中，目前关于“世界模型”研究的主要进展集中在“空间认知”上。在这方面，1974年伦敦大学学院的约翰·奥基夫（John O’Keefe）教授在大脑海马区发现了位置神经元^[14]。位置神经元能够在不同的空间位置上表现出一定的位置感受野。1990年，塔布（Taube）发现了位于丘脑后下区的头朝向神经元，其放电活动在不同的头朝向上表现

出一定的感受野^[83]。2005年，挪威神经科学家莫泽夫妇（May-Britt Moser 和 Edvard Moser）在大脑内嗅皮层发现了网格细胞。网格细胞也表现出一定的空间相关性。与位置神经元不同的是，网格细胞的感受野呈现规律的网格状，且在不同的环境中表现稳定^[84]。此后，科学家们还陆续发现了边界细胞^[85]、速度细胞^[86]。这些研究引起了科学家们极大的兴趣，并提出了很多的计算模型，以解释大脑“空间认知”的工作机理（例如方向感、空间导航等问题^[87-88]），并试图向“非空间认知”拓展，形成完整的“世界模型”^[89]。但目前这方面并未形成一个完整的结论。

如前所述，“世界模型”表示了一切外部事物之间的关系。因此，“世界模型”表示的范围非常广泛。这绝非式（17）和式（18）所示的简单模型能表示的。事实上，式（17）和式（18）的表示是有很大的缺陷的。例如，式（17）和式（18）表示了 s_{n+1} 和 s_n 之间的关系，却没有表示 s_{n+1} 和 $s_{n-1}, s_{n-2}, \dots$ 之间的关系。这对数据的利用是不充分的。因此，“世界模型”的表征问题存在很大的研究空间。在前面基于值迭代的强化学习中，人们随机初始化 $V(s)$ 。在机器人学科中，往往利用先验知识设计一个 $V(s)$ 的初始值，从而加速机器人的学习进程。这

个初始化的 $V(s)$ 被称为回报塑形(reward shaping)^[90], 可认为其是“世界模型”的一种表征。“知识”也同样属于“世界模型”的范畴。在计算机学科中, 随着互联网技术的发展, “知识表示”受到了空前的重视, 人们提出了“知识图谱”的概念^[91]。“知识图谱”“知识表示”同样也可以认为是“世界模型”的一部分(或一种表示)。“世界模型”的表征问题是大脑研究中非常核心的一个问题, 这还有赖于多学科共同努力。

7 大脑建模研究中的几个热点问题

7.1 热点问题

从上文可以看到, 强化学习是大脑建模中非常核心的部分, 也是大脑建模中十分精彩的部分。以强化学习为核心的大脑模型已经能够描述图7的分布式自适应控制模型的主体。但是, 这里仍然有很多问题尚未解决。

7.1.1 记忆是如何参与决策的?

2018年, 英国 DeepMind 公司发表了一篇题为《在这里, 那样做: 基于情景回忆的元学习 (been there, done that: meta-learning with episodic recall)》的文章^[92], 提出了简单的记忆以及记忆如何参与决策过程的模型, 如图18所示。但是, 这篇文章中所提的方法只是一个很好的“机器学习”算法, 并无太多的神经科学和心理学依据。

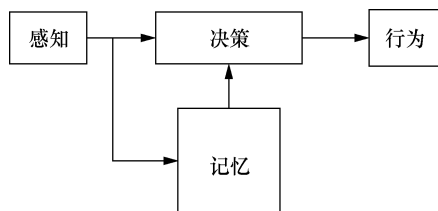


图18 记忆参与决策的模型

如前所述, 记忆既包括当前外部客观世界在大脑中的反映, 也包括过去时间外部客观世界在大脑中的反映, 甚至还包括过去时间大脑的意识活动(即曾经“思考”过什么问题)在大脑中的反映。在基于模型的强化学习过程中, 同时会建立一个“世界模型”, 这可以说是记忆的初步模型。但是, 该模型远不是一个简单的前向神经网络所能表达的, 它在大脑中具体是怎样建立的以及大脑具体如何使用该模型进行决策, 仍然是一个悬而未决的问题。基于模型的强化学习和即时定位与地图构建(simultaneous localization and mapping, SLAM)^[93-94]

算法为人们提供了一个很好的框架和基础, 但问题依然摆在人们面前。

7.1.2 部分观测马尔可夫模型

2016年, DeepMind 公司发布了 AlphaGo 算法。采用该算法的计算机在围棋博弈中, 具有超越人类的能力, 打败了国际围棋冠军李世石。在每一步的对弈中, 该算法对棋盘状况具有完全的知识, 但是在人类活动的很多场合, 例如市场经济中的公司决策、股票市场决策、战场决策等, 人脑面临非完全信息, 而且必须在此情形下做出决策。扑克牌游戏就属于典型的非完备信息条件下的决策问题^[43]。2015年, 加拿大阿尔伯塔大学提出了一对一的德克萨斯扑克游戏算法, 解决了一对一的非完备信息条件下的决策问题^[95]。在更广泛条件下, 人脑在非完备信息条件下做出决策的机制尚不明晰; 目前也没有一种机器学习算法能达到人类专家在此情形下的决策能力。这是当前业界面临的一个重大问题。

7.1.3 基于模型的规划问题

当人们有了一个目标, 人们需要选择一个策略以达到该目标。策略的选择就是规划问题。规划问题既可以是空间路径的规划问题, 也可以是其他非空间目标的实现策略规划问题。当人们的大脑中有一个“世界模型”时, 即便是对于未曾到达过的目标, 人们一般也可以很快地生成一个策略。但是, 目前并不清楚该策略到底是如何生成的。有学者提出可以用生成神经网络来生成该策略^[96-97]。将当前状态和目标状态作为生成神经网络的输入, 生成神经网络将输出一个策略。但是, 这里依旧存在两个问题: 第一, 由于目标和策略的多样性, 该生成神经网络将十分庞大; 第二, 这样一个庞大的生成神经网络需要大量的训练。这对人们搞清楚这个生成神经网络的结构是不利的。因此, 简单地说一个生成神经网络仍然太笼统。这个问题仍然需要进一步研究。

7.1.4 问题分解模型

人类在思考问题时, 往往会把一个大问题分解为若干个小问题, 然后把小问题分解为更小的问题, 如此分层次划分, 对问题进行树状分解。树的叶节点都是很容易解决的细小问题。那么, 问题分解的标准是什么? 如何用程序描述这个过程? 这些都是悬而未决的问题。目前科学界提出了几种模型: 参考文献[98]提出分层强化学习模型; 参考文献[99]提出不同时间尺度的并行时间差分模型; 参

考文献[100]提出“选择(options)”的概念,并提出人脑发现 options 的模型。但是,这些模型在解决不同问题时,仍然需要较多的人工参与和尝试,它们对问题分解的通用性仍然差强人意。这是目前人工智能和人脑工作记忆领域的一个核心问题。

7.1.5 需求如何产生动机,动机如何产生目标

如前所述,人的生理需求会驱动内感觉系统,产生动机,动机进一步产生目标,目标指导行为。动机可认为是潜意识的、深层次的潜在目标。与动机相比,目标具有显性和浅层次的特点^[101]。现代神经科学研究表明,大脑的边缘系统(包括下丘脑、海马和杏仁核)在人的动机和情感生成中扮演重要角色。人的大脑每时每刻都会有很多个潜在的目标。大脑会比较权衡实现各个目标的代价和收益,从而选取其中一个目标来指导行为。为了对各个潜在目标的代价和收益进行比较,大脑需要一个统一的货币进行评估^[102]。

但是,生理需求究竟如何产生动机?动机如何产生潜在目标?大脑如何建立统一货币?大脑又是如何使用该统一货币对各个不同潜在目标的代价和收益进行定价的?这些问题仍然没有搞明白。这对很多大脑疾病的治疗(例如精神失常(psychiatric disorder))和实现通用人工智能都非常重要。

7.1.6 改变主意的机理

第7.1.5节从动机产生目标的角度说明了大脑有可能突然改变主意,选取不同的目标,但原目标并没有消失。还有一种改变主意的机理,即大脑根据某个目标做出一个初始决策后,在按照该决策进行的过程中,发现情况与原先估计有较大差别,于是放弃该目标。这说明在初始决策时,大脑对决策后的未来状态有一个估计,并且伴随该估计,通常有一个估计的信心^[27-29,52,73,103],在初始决策后的继续行进过程中,大脑会进一步地使用当前的观察输入来印证先前的估计,这叫作证据累积(evidence accumulation)。当证据累积达到相反的门槛时(即累积误差超过一定门限时),大脑做出改变主意的决策^[104-105]。

以上只是形象化地描述了大脑改变主意的过程。大脑改变主意的确切机理和过程仍然不清楚。这是当前研究的又一个热点问题。

7.2 与相关学科的关系

7.2.1 与机器人研究的关系

熟悉“机器人学”的研究人员会发现,图5(b)

及图6是典型的现代机器人理论架构^[106-107]。事实上,“机器人”的目标就是用“机器”替代“人”,这就要求“机器行为”能够尽量复制“人类行为”,这与图1中要求“大脑模型行为”与“大脑行为”相一致是相同的。因此,在大脑建模中,可以借鉴和参考大量的“机器人学”理论,作为“猜测”大脑模型的出发点。不同的是,机器人研究只要求模型具有可实现性,不要求模型的具体实现形式;而大脑建模研究,不但要求模型具有可实现性,还要求模型尽量符合大脑的生物物理原理,这是比机器人研究更严苛的要求。也就是说,机器人研究具有比大脑建模研究更大的灵活空间。一方面,从机器人研究出发,逐渐下降抽象层次,转换为大脑建模研究,不失为一种好的方法;另一方面,大脑研究积累了很多宏观的结构和行为知识,这些生物的结构和行为知识也可以启发人们对机器人的研究^[108]。例如,机器人研究中的“性能监视(performance monitor)”模块就是受大脑研究中的“元认知”启发得来的^[28-29,52,103,109-110]。因此,总体上来说,二者联系非常紧密,相辅相成。

7.2.2 与数据科学的关系

大脑的重要功能是具有学习能力。所谓学习,就是根据过去的经验来改善自身技能和产生认知的过程^[111]。从数据科学的角度来说,学习就是根据输入/观测数据改变模型参数和状态的过程。例如,式(17)和式(18)就是在深入挖掘观测数据之间关系的基础上提出的。数据科学就是主要研究数据之间关系和数据处理方法的学科,它不受任何模型形式和研究目标的约束,具有比大脑建模研究、机器人研究更广泛、更灵活的研究空间。而且,数据科学研究往往不只追求实用性,还追求理论上的最优化。一方面,数据科学研究可以为大脑研究提供更加根本、更加基础的概念和理论。这尤其表现在大脑的感知推理(perceptual inference)当中^[112]。另一方面,大脑研究亦将对数据科学有所裨益。例如,大脑的“探索”“操作性学习”等行为可以为数据科学中新数据的收集提供一定的参考;大脑内“世界模型”或“知识”的表达,极可能引起数据组织方式的变革。因此,总体上来说,二者互有裨益;但短期来看,主要是大脑研究受益于数据科学研究。

但是需要强调的是,机器人、机器学习和大数据理论只能给大脑建模研究提供一定的概念灵感

和理论基础, 并不能最终解决大脑工作机理的问题。例如, 猴子和人可以经历相同的观察学习, 人脑能学会, 而猴脑不能学会。因此, 大脑工作机理的更深层问题还有赖于神经生物学的研究和发现。

8 结束语

理解大脑的工作原理, 是目前全人类共同面临的一个重大课题。它一方面将为人类创造“终极”的人工智能奠定基础; 另一方面, 将为人类大脑疾病的诊断和治疗做好铺垫。从 20 世纪末期以来, 人类已经在理解大脑的工作原理方面取得了相当多的进展, 已经发展成为名为“计算神经科学/计算认知科学”的学科^[113]。甚至有学者宣称, “计算神经科学”已经进入“中年期”^[114]。但是, 从本文的介绍可以看到, 大脑建模仍然有很多的问题尚未解决。而且, 即便是已经提出的一些模型, 它们的生物物理基础也并不十分牢靠。因此, 说“计算神经科学”尚处于“少年期”或许更合适。但不管怎样, 已有的工作给研究者们打下了坚实的基础, 成功的曙光已经初现端倪, 这将激励着这一代的科学家们不断向前!

参考文献:

- [1] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [2] 钟义信. 机制主义人工智能理论: 一种通用的人工智能理论[J]. *智能系统学报*, 2018, 13(1): 2-18.
ZHONG Y X. Mechanism-based artificial intelligence theory: a universal theory of artificial intelligence[J]. *CAAI Transactions on Intelligent Systems*, 2018, 13(1): 2-18.
- [3] 何华灿. 泛逻辑学理论: 机制主义人工智能理论的逻辑基础[J]. *智能系统学报*, 2018, 13(1): 19-36.
HE H C. Universal logic theory: logical foundation of mechanism-based artificial intelligence theory[J]. *CAAI Transactions on Intelligent Systems*, 2018, 13(1): 19-36.
- [4] 汪培庄. 因素空间理论: 机制主义人工智能理论的数学基础[J]. *智能系统学报*, 2018, 13(1): 37-54.
WANG P Z. Factor space-mathematical basis of mechanism based artificial intelligence theory[J]. *CAAI Transactions on Intelligent Systems*, 2018, 13(1): 37-54.
- [5] 汪培庄, 周红军, 何华灿, 等. 因素表示的信息空间与广义概率逻辑[J]. *智能系统学报*, 2019, 14(5): 843-852.
WANG P Z, ZHOU H J, HE H C, et al. Factorial information space and generalized probability logic[J]. *CAAI Transactions on Intelligent Systems*, 2019, 14(5): 843-852.
- [6] 刘经纬, 赵辉, 周瑞, 等. 规则推理与神经计算智能控制系统改进及比较[J]. *智能系统学报*, 2017, 12(6): 823-832.
LIU J W, ZHAO H, ZHOU R, et al. Improvement and comparison research between intelligent control systems based on rule based reasoning and neural computation AI methods[J]. *CAAI Transactions on Intelligent Systems*, 2017, 12(6): 823-832.
- [7] 李德毅. AI: 人类社会发展的加速器[J]. *智能系统学报*, 2017, 12(5): 583-589.
LI D Y. Artificial intelligence: an accelerator for the development of human society[J]. *CAAI Transactions on Intelligent Systems*, 2017, 12(5): 583-589.
- [8] 曾毅, 刘成林, 谭铁牛. 类脑智能研究的回顾与展望[J]. *计算机学报*, 2016, 39(1): 212-222.
ZENG Y, LIU C L, TAN T N. Retrospect and outlook of brain-inspired intelligence research[J]. *Chinese Journal of Computers*, 2016, 39(1): 212-222.
- [9] FASANO A, DANIELE A, ALBANESE A. Treatment of motor and non-motor features of Parkinson's disease with deep brain stimulation[J]. *The Lancet Neurology*, 2012, 11(5): 429-442.
- [10] KUO C C J. Understanding convolutional neural networks with a mathematical model[J]. *Journal of Visual Communication and Image Representation*, 2016, 41: 406-413.
- [11] 寿天德. 神经生物学[M]. 3 版. 北京: 高等教育出版社, 2013.
SHOU T D. *Neurobiology*[M]. 3rd edition. Beijing: Higher Education Press, 2013.
- [12] KANDEL E, SCHWARTZ J, JESSELL T, et al. *Principles of neural science*[M]. 5th edition. [S.l.:s.n.], 2013.
- [13] DAYAN P, ABBOTT L F. *Theoretical neuroscience*[M]. Cambridge: The MIT Press, 2001.
- [14] ECKARDT M. The hippocampus as a cognitive map[J]. *The Journal of Nervous and Mental Disease*, 1980, 168(3): 191-192.
- [15] LI Z P. *Understanding vision: theory, models, and data*[M]. Oxford: Oxford University Press, 2014.
- [16] GEORGIEVA P, SILVA F, MILANOVA M, et al. EEG signal processing for brain-computer interfaces[M]//*Springer handbook of bio-/neuroinformatics*. Heidelberg: Springer, 2014.
- [17] LIU Z Q. Causation, Bayesian networks, and cognitive MapsFull-text in English[J]. *Acta Automatica Sinica*, 2001, 27(4): 552-566.
- [18] BARBER D. *Machine learning*[M]//*Bayesian reasoning and machine learning*. Cambridge: Cambridge University Press, 2012: 303-304.
- [19] AXELROD R. *Structure of decision: the cognitive maps of political elites*[M]. Princeton: Princeton University Press, 2016.
- [20] FREDRIC M H, IVICA K. *神经计算原理*[M]. 叶世伟, 王海娟, 译. 北京: 机械工业出版社, 2007.
FREDRIC M H, IVIC K. *Principles of neurocomputing for science & engineering*[M]. Translated by YE S W, WANG H J. Beijing: China Machine Press, 2007.
- [21] DE KAMPS M, VAN DER VELDE F. From artificial neural networks to spiking neuron populations and back again[J]. *Neural Networks*, 2001, 14(6/7): 941-953.
- [22] FURBER S. *I origins*[M]//*SpiNNaker: a spiking neural network architecture*. Boston-Delft: Now Publishers, 2020.
- [23] GERSTNER W, KISTLER W M, NAUD R, et al. *Neuronal dynamics*[M]. Cambridge: Cambridge University Press, 2009.
- [24] BRETTE R. Computing with neural synchrony[J]. *PLoS Computational Biology*, 2012, 8(6): e1002561.
- [25] VERSCHURE P F M J, PENNARTZ C M A, PEZZULO G. The why, what, where, when and how of goal-directed choice: neuronal and computational principles[J]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2014, 369(1655): 20130483.

- [26] VERSCHURE PAUL F M J. Distributed adaptive control: a theory of the mind, brain, body nexus[J]. *Biologically Inspired Cognitive Architectures*, 2012: 55-72.
- [27] BOLDT A, BLUNDELL C, DE MARTINO B. Confidence modulates exploration and exploitation in value-based learning[J]. *Neuroscience of Consciousness*, 2019, 2019(1): niz004.
- [28] VEENMAN M V J, VAN HOUT-WOLTERS B H A M, AFFLERBACH P. Metacognition and learning: conceptual and methodological considerations[J]. *Metacognition and Learning*, 2006, 1(1): 3-14.
- [29] COX M T. Metacognition in computation: a selected research review[J]. *Artificial Intelligence*, 2005, 169(2): 104-141.
- [30] MAASS W. On the computational power of winner-take-all[J]. *Neural Computation*, 2000, 12(11): 2519-2535.
- [31] FARZANEH Y, AKBARZADEH A, AKBARI A A. New automated learning CPG for rhythmic patterns[J]. *Intelligent Service Robotics*, 2012, 5(3): 169-177.
- [32] GUERTIN P A. Central pattern generator for locomotion: anatomical, physiological, and pathophysiological considerations[J]. *Frontiers in Neurology*, 2012, 3: 183.
- [33] KATZ P S. Evolution of central pattern generators and rhythmic behaviours[J]. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 2016, 371(1685): 20150057.
- [34] KARL F. The free-energy principle: a rough guide to the brain?[J]. *Trends in Cognitive Sciences*, 2009, 13(7): 293-301.
- [35] BOGACZ R. A tutorial on the free-energy framework for modelling perception and learning[J]. *Journal of Mathematical Psychology*, 2017, 76: 198-211.
- [36] KINGMA D P, WELING M. An introduction to variational autoencoders[J]. *Foundations and Trends® in Machine Learning*, 2019, 12(4): 307-392.
- [37] HINTON G. Boltzmann machine[J]. *Scholarpedia*, 2007, 2(5): 1668.
- [38] SALAKHUTDINOV R, HINTON G E. Deep Boltzmann machines[J]. *Journal of Machine Learning Research*, 2009, 5(2): 448-455.
- [39] BACHMAN P, SORDONI A, TRISCHLER A. Learning algorithms for active learning[C]//*Proceedings of the 34th International Conference on Machine Learning*. [S.l.:s.n.], 2017.
- [40] DOUGLAS R J, MARTIN K A C. Recurrent neuronal circuits in the neocortex[J]. *Current Biology: CB*, 2007, 17(13): 496-500.
- [41] WANG X J. Decision making in recurrent neuronal circuits[J]. *Neuron*, 2008, 60(2): 215-234.
- [42] MORSE A F, DE GREEFF J, BELPEAME T, et al. Epigenetic robotics architecture (ERA)[J]. *IEEE Transactions on Autonomous Mental Development*, 2010, 2(4): 325-339.
- [43] DAYAN P, DAW N D. Decision theory, reinforcement learning, and the brain[J]. *Cognitive, Affective & Behavioral Neuroscience*, 2008, 8(4): 429-453.
- [44] DABNEY W, KURTH-NELSON Z, UCHIDA N, et al. A distributional code for value in dopamine-based reinforcement learning[J]. *Nature*, 2020, 577(7792): 671-675.
- [45] MILLER K J, SHENHAV A, LUDVIG E A. Habits without values[J]. *Psychological Review*, 2019, 126(2): 292-311.
- [46] DOLAN R J, DAYAN P. Goals and habits in the brain[J]. *Neuron*, 2013, 80(2): 312-325.
- [47] FRIEDEL E, KOCH S P, WENDT J, et al. Devaluation and sequential decisions: linking goal-directed and model-based behavior[J]. *Frontiers in Human Neuroscience*, 2014, 8: 587.
- [48] PEZZULO G, VERSCHURE P F M J, BALKENIUS C, et al. The principles of goal-directed decision-making: from neural mechanisms to computation and robotics[J]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2014, 369(1655): 20130470.
- [49] DAW N D, DAYAN P. The algorithmic anatomy of model-based evaluation[J]. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 2014, 369(1655): 20130478.
- [50] BUSCHMAN T J, MILLER E K. Goal-direction and top-down control[J]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2014, 369(1655): 20130471.
- [51] BOTVINICK M, WEINSTEIN A. Model-based hierarchical reinforcement learning and human action control[J]. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2014, 369(1655): 20130480.
- [52] BOUREAU Y L, SOKOL-HESSNER P, DAW N D. Deciding how to decide: self-control and meta-decision making[J]. *Trends in Cognitive Sciences*, 2015, 19(11): 700-710.
- [53] GLIMCHER P W, FEHR E. *Neuroeconomics: decision making and the brain*[M]. 2nd edition. [S.l.:s.n.], 2014.
- [54] HA D, SCHMIDHUBER J. World models[J]. *arXiv preprint*, 2018, arXiv:1803.10122.
- [55] FU M C. Monte Carlo tree search: a tutorial[C]//*Proceedings of 2018 Winter Simulation Conference (WSC)*. Piscataway: IEEE Press, 2018: 222-236.
- [56] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. *Nature*, 2016, 529(7587): 484-489.
- [57] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge[J]. *Nature*, 2017, 550(7676): 354-359.
- [58] MORAN R, KERAMATI M, DAYAN P, et al. Retrospective model-based inference guides model-free credit assignment[J]. *Nature Communications*, 2019, 10: 750.
- [59] GERSHMAN S J, MARKMAN A B, OTTO A R. Retrospective revaluation in sequential decision making: a tale of two systems[J]. *Journal of Experimental Psychology General*, 2014, 143(1): 182-194.
- [60] KRÄMER W. Kahneman, D. (2011): thinking, fast and slow[J]. *Statistical Papers*, 2014, 55(3): 915.
- [61] BANSAL S, CALANDRA R, CHUA K, et al. MBMF: model-based priors for model-free reinforcement learning[J]. *arXiv preprint*, 2017, arXiv:1709.03153.
- [62] DEISENROTH M P, RASMUSSEN C E. PILCO: a model-based and data-efficient approach to policy search[C]//*Proceedings of the 28th International Conference on Machine Learning*. [S.l.:s.n.], 2011: 465-472.
- [63] SMITTENAAR P, FITZGERALD T H B, ROMEI V, et al. Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans[J]. *Neuron*, 2013, 80(4): 914-919.
- [64] KERAMATI M, DEZFOULI A, PIRAY P. Speed/accuracy trade-off between the habitual and the goal-directed processes[J]. *PLoS Computational Biology*, 2011, 7(5): e1002055.
- [65] KOEHLIN E, SUMMERFIELD C. An information theoretical approach to prefrontal executive function[J]. *Trends in Cognitive Sciences*, 2007, 11(6): 229-235.
- [66] TYE K M, JANAK P H. Amygdala neurons differentially encode

- motivation and reinforcement[J]. The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 2007, 27(15): 3937-3945.
- [67] SCHÖNFELD L M, ZECH M P, SCHÄBLE S, et al. Lesions of the rat basolateral amygdala reduce the behavioral response to ultrasonic vocalizations[J]. Behavioural Brain Research, 2020, 378: 112274.
- [68] PANKSEPP J. Affective neuroscience: the foundations of human and animal emotions[M]. Oxford: Oxford University Press, 2005.
- [69] SMITH R, ALKOZEI A, KILLGORE W D S. How do emotions work?[J]. Frontiers for Young Minds, 2017, 5: 69.
- [70] 于冬梅. 情感计算关键技术研究[D]. 上海: 东华大学, 2009.
- YU D M. Research on key technology of affective computing[D]. Shanghai: Donghua University, 2009.
- [71] VALIKHANI M, SOURKOUNIS C. A brain emotional learning-based intelligent controller (BELBIC) for DFIG system[C]//Proceedings of 2014 International Symposium on Power Electronics, Electrical Drives, Automation and Motion. Piscataway: IEEE Press, 2014: 713-718.
- [72] MCDUFF D, KAPOOR A. Visceral machines: risk-aversion in reinforcement learning with intrinsic physiological rewards[C]//Proceedings of 2019 Conference Blind Submission. [S.l.:s.n.], 2019.
- [73] NAVAJAS J, HINDOCHA C, FODA H, et al. The idiosyncratic nature of confidence[J]. Nature Human Behaviour, 2017, 1(11): 810-818.
- [74] BISHOP S J, GAGNE C. Anxiety, depression, and decision making: a computational perspective[J]. Annual Review of Neuroscience, 2018, 41: 371-388.
- [75] LERNER J S, LI Y, VALDESOLO P, et al. Emotion and decision making[J]. Annual Review of Psychology, 2015, 66(1): 799-823.
- [76] OTTER D W, MEDINA J R, KALITA J K. A survey of the usages of deep learning for natural language processing[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(2): 604-624.
- [77] 朱张莉, 饶元, 吴洲, 等. 注意力机制在深度学习中的研究进展[J]. 中文信息学报, 2019, 33(6): 1-11.
- ZHU Z L, RAO Y, WU Y, et al. Research progress of attention mechanism in deep learning[J]. Journal of Chinese Information Processing, 2019, 33(6): 1-11.
- [78] ANTUNES A, LAFLAQUIERE A, OGATA T, et al. A Bi-directional multiple timescales LSTM model for grounding of actions and verbs[C]//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press, 2019: 2614-2621.
- [79] ŠTEPÁNOVÁ K, KLEIN F B, CANGELOSI A, et al. Mapping language to vision in a real-world robotic scenario[J]. IEEE Transactions on Cognitive and Developmental Systems, 2018, 10(3): 784-794.
- [80] FELDMAN H, FRISTON K J. Attention, uncertainty, and free-energy[J]. Frontiers in Human Neuroscience, 2010, 4: 215.
- [81] STACHENFELD K L, BOTVINICK M M, GERSHMAN S J. The hippocampus as a predictive map[J]. Nature Neuroscience, 2017, 20(11): 1643-1653.
- [82] LAND M F. Do we have an internal model of the outside world?[J]. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 2014, 369(1636): 20130045.
- [83] WINTER S S, TAUBE J S. Head direction cells: from generation to integration[M]//Space, time and memory in the hippocampal formation. Vienna: Springer Vienna, 2014: 83-106.
- [84] ROWLAND D C, ROUDI Y, MOSER M B, et al. Ten years of grid cells[J]. Annual Review of Neuroscience, 2016, 39: 19-40.
- [85] BARRY C, LEVER C, HAYMAN R, et al. The boundary vector cell model of place cell firing and spatial memory[J]. Reviews in the Neurosciences, 2006, 17(1/2): 71-97.
- [86] KROPFF E, CARMICHAEL J E, MOSER M B, et al. Speed cells in the medial entorhinal cortex[J]. Nature, 2015, 523(7561): 419-424.
- [87] CULLEN K E, TAUBE J S. Our sense of direction: progress, controversies and challenges[J]. Nature Neuroscience, 2017, 20(11): 1465-1473.
- [88] PARK E H, KEELEY S, SAVIN C, et al. How the internally organized direction sense is used to navigate[J]. Neuron, 2019, 101(2): 285-293.
- [89] BEHRENS T E J, MULLER T H, WHITTINGTON J C R, et al. What is a cognitive map? Organizing knowledge for flexible behavior[J]. Neuron, 2018, 100(2): 490-509.
- [90] KOBER J, BAGNELL J A, PETERS J. Reinforcement learning in robotics: a survey[J]. The International Journal of Robotics Research, 2013, 32(11): 1238-1274.
- [91] DU Z X, ZHOU C, DING M, et al. Cognitive knowledge graph reasoning for one-shot relational learning[J]. arXiv preprint, 2019, arXiv:1906.05489.
- [92] RITTER S, WANG J X, KURTH-NELSON Z, et al. Been there, done that: meta-learning with episodic recall[C]//Proceedings of the 35th International Conference on Machine Learning. [S.l.:s.n.], 2018.
- [93] DURRANT-WHYTE H, BAILEY T. Simultaneous localization and mapping: part I[J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99-110.
- [94] BAILEY T, DURRANT-WHYTE H. Simultaneous localization and mapping (SLAM): part II[J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108-117.
- [95] BOWLING M, BURCH N, JOHANSON M, et al. Heads-up limit hold'em poker is solved[J]. Science, 2015, 347(6218): 145-149.
- [96] FAULKNER R, PRECUP D. Dyna planning using a feature based generative model[J]. arXiv preprint, 2018, arXiv:1805.10129.
- [97] SOLWAY A, BOTVINICK M M. Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates[J]. Psychological Review, 2012, 119(1): 120-154.
- [98] BOTVINICK M M, NIV Y, BARTO A G. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective[J]. Cognition, 2009, 113(3): 262-280.
- [99] PONG V, GU S X, DALAL M, et al. Temporal difference models: model-free deep RL for model-based control[J]. arXiv preprint, 2018: arXiv:1802.09081.
- [100] MAISTO D, DONNARUMMA F, PEZZULO G. Divide et impera: subgoalng reduces the complexity of probabilistic inference and problem solving[J]. Journal of the Royal Society Interface, 2015, 12(104): 20141335.
- [101] BERKMAN E T. The neuroscience of goals and behavior change[J]. Consulting Psychology Journal: Practice and Research, 2018, 70(1): 28-44.
- [102] KOOL W, BOTVINICK M. Mental labour[J]. Nature Human Behaviour, 2018, 2(12): 899-908.
- [103] FLEMING S M, DAW N D. Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation[J]. Psychological Review, 2017, 124(1): 91-114.
- [104] NAVAJAS J, BAHRAMI B, LATHAM P E. Post-decisional accounts of biases in confidence[J]. Current Opinion in Behavioral Sciences,

- 2016, 11: 55-60.
- [105]ATTIYA N A A, RAÑÓ I, PRASAD G, et al. A neural circuit model of decision uncertainty and change-of-mind[J]. Nature Communications, 2019, 10: 2287.
- [106]MURPHY R R. Introduction to AI robotics[J]. Industrial Robot: an International Journal, 2001, 28(3): 266-267.
- [107]DE SILVA L, EKANAYAKE H. Behavior-based robotics and the reactive paradigm a survey[C]//Proceedings of 2008 11th International Conference on Computer and Information Technology. Piscataway: IEEE Press, 2008: 36-43.
- [108]HASSABIS D, KUMARAN D, SUMMERFIELD C, et al. Neuroscience-inspired artificial intelligence[J]. Neuron, 2017, 95(2): 245-258.
- [109]KRALIK J D, MAO T, CHENG Z, et al. Modeling incubation and restructuring for creative problem solving in robots[J]. Robotics and Autonomous Systems, 2016, 86: 162-173.
- [110]WINFIELD A F T. Experiments in artificial theory of mind: from safety to story-telling[J]. Frontiers in Robotics and AI, 2018, 5: 75.
- [111]GROSS R. Psychology: the science of mind and behavior (6E)[M]. UK: Hodder Education, 2010.
- [112]SHAMS L, BEIERHOLM U R. Causal inference in perception[J]. Trends in Cognitive Sciences, 2010, 14(9): 425-432.
- [113]MICHEL M, BECK D, BLOCK N, et al. Opportunities and challenges for a maturing science of consciousness[J]. Nature Human Behaviour,

2019, 3(2): 104-107.

- [114]DAYAN P, DURSTEWITZ D. Editorial overview: computational modeling: after the middle ages of computational neuroscience[J]. Current Opinion in Behavioral Sciences, 2016, 11: v-vi.

[作者简介]



胡东伟（1980—），男，博士，中国电子科技集团公司第五十四研究所高级工程师，主要研究方向为无线通信理论、大规模集成电路设计和计算神经科学。



冯晓璐（1988—），女，石家庄二中实验学校教师，主要研究方向为教育心理学和语言学。