JOURNAL OF EAST CHINA NORMAL UNIVERSITY
Humanities and Social Sciences No.5, 2018

具身性与默会表征:人工智能能走多远?

何 静

摘 要:人类智能的探究和实现问题既是一个哲学问题,也是一个科学问题。人类智能从本质上说是具身的,但同时也紧密地依赖于计算和表征的大脑。导致人工智能研究困境的根源并非在于它采用了计算—表征的框架,而在于它没能将关于身体的信息涵盖进它对智能程序的描述中去。因此,承载着大量实践性知识的默会表征对未来人工智能研究在拓展解决问题常规以及降低计算复杂性等方面有着重要意义。

关键词:人工智能;计算;具身性;默会表征;实践性知识

DOI: 10.16382/j.cnki.1000-5579.2018.05.005

一 导 言

认知科学是哲学认识论的延续,是研究者们为了回答认识论问题所作出的经验性的尝试。它关注人类的知识结构与智能实现,同时也因为一些新出现的经验主题而得到扩展。几十年来,认知科学经历了从行为主义到符号主义,再到联结主义的范式转换;经历了从最初的问题求解到人工神经网络的研究,从符号计算到多层神经网络动态交互的实施。这一系列的范式转换,体现了研究者们不断吸收科学实验对心智现象的研究成果,并对其研究纲领进行局部修正以求新突破的不断尝试。随着学科建制化步伐的加速,认知科学已经成为一门独立的学科,形成了相对稳定的理论框架和研究范式。而人工智能也已成为这一学科的智力内核,取得了举世瞩目的成就。

然而,尽管今天的人工智能能够完成人类无法完成的复杂计算,做人类不能做的精密工作,却也遇到了各种理论困境和实践困难。[①]这些技术困境背后其实隐藏着更深层次的认识论问题:一方面,人类的智力过程被认为是通过大脑来完成和实现的;另一方面,智能与世界之间的关联被假定为一系列抽象的表征(representations)。这种将大脑作为推理机的假设具有误导性。因为人类在真实世界中的问题解决能力似乎并不总来自大脑中明确的数据结构,而常常以一种难以预期的方式从身体与世界打交道的过程中涌现出来。换言之,人类智能是具身行动中熟练的实践性技能的运用(know-how)。这意味着,一旦远离了身体以及身体与真实世界的勾连,人工智能系统便很难对直接利用身体能力与世界结构来简化或改变认知方式的人类智能进行刻画。正如谢菲尔和斯凯耶尔(R.Pfeifer & C.Scheier)所说:"智能不仅仅是一种抽象的运算法则的形式。它需要身体的示例(physical instantiation)和肉体的介入。"[②]

如果身体的属性确实"塑造"了人类的智能特征,那么人工智能学家就应该把身体的相关属性表征纳入构成认知的算法中。因此,未来人工智能研究的核心关切将不仅仅是对人类"智能"的模拟,还将包括对"身体"的设计。本文通过考察克拉克(A. Clark)对蟑螂逃生技能的探究以及丹尼特(D.

【作者简介】何静,哲学博士,华东师范大学哲学系副教授(上海,200241)。

【基金项目】国家社科基金项目"具身哲学视域中的社会认知研究"(项目编号:15CZX015)。

① 参见刘晓力:《认知科学研究纲领的困境与走向》,《中国社会科学》2003年第1期;Larry Shapiro, *Embodied Cognition*, New York: Routledge。

② Rolf Pfeifer & Christian Scheier, *Understanding Intelligence*, MIT Press, 1999, p.23.

Dennett)对明确表征(explicit representation)和默会表征(tacit representation)的阐释,质疑了传统表征概念在智能解释中的作用,并进一步揭示具身性与默会表征对未来人工智能研究在拓展解决问题常规以及降低计算复杂性方面的重要意义。

二 符号表征与孤立的大脑

对身体的鄙薄,是自柏拉图以来西方思想的一贯传统。到了17世纪的笛卡尔那里,这一倾向得到了进一步强化。笛卡尔的身心二元论强调"旁观者"的知识模式,以绝对客观性为知识的理想,视心智为自然之镜,视认知为对外部世界的表征。这种发端于笛卡尔,中经洛克和康德的客观主义认识论,在19世纪和20世纪的整个哲学界得到延续。客观主义认识论以表征主义为特征,预设了主体与客体、内部心智与外部世界的分离;身体没有在对概念意义的表述中扮演实质性的角色,也没有在描述推理本质的过程中发挥关键作用。

这种客观主义的哲学传统为当代人工智能的发展提供了理论基础。研究者们主张:人类的智力过程就是通过大脑对符号表征的产生、转化和操作来完成的。[①]首先,大脑将非符号的知觉输入转换为任务域的符号表征,而后以一种纯句法的形式对这些表征进行处理,并最终形成解决问题的方案。由此,研究者们强调,人工智能研究应当关注对认知任务(即符号表征的结构和内容)的抽象描述以及为解决问题而对表征进行操作的算法本质。在这个方面,经典人工智能与功能主义(functionalism)同流,认为人类的智能在本质上与具身性无关;对于所有的智能形式而言(无论是自然的,还是人工的),至关重要的是软件而不是硬件。

在这种计算强纲领的指引下,早期的人工智能研究者们进行了深入的研究。其中以纽维尔和西蒙(A. Newell & H. Simon)的通用求解器研究最具代表性。纽维尔和西蒙提出了"物理符号假设"(PSSH),即"一个物理符号系统既是通用智能行为的充分条件,也是必要条件"[②],并设计了一种叫作"通用问题求解器"(GPS)的计算机程序。GPS的目标不仅仅是要解决逻辑问题,而且要模拟人类在解决现实世界逻辑问题时的脑部活动。纽维尔和西蒙通过将逻辑公式转换为"意义—目标分析"(means-end analysis),并用关于初始状态、目标状态以及一组从初始状态到目标状态路径规则的表征来解决特定领域中的问题。到1957年,GPS已经能够成功地下国际象棋、证明几何定理以及设计机器程序。这是人工智能研究的先驱者们试图通过基于规则的符号操作,在形式系统中模拟人类智能的伟大尝试。但是,GPS也遭遇了两个似乎无法逾越的难题:有限的实用性以及有限的运用范围。这些都与GPS当年预期的"通用思维模拟"梦想相去甚远。

在此后的几十年间,人工智能研究试图摆脱"符号—表征"范式的困境,而通过人工神经网络来模拟大脑非线性和分布式计算的特征。人工神经网络将类神经元的节点通过数值加权的联结方式连接起来,并通过联结权重的调节来适应学习规则和系统活动的变化。当代人工智能研究中最令人兴奋、最具开拓性的研究成果就是深度学习(deep learning)技术。这种技术以神经生物学的大脑为原型,用自治体的方法论来提升机器的智能。它将智能解释的重点放在神经网络的结构、学习规则和从多层神经网络间涌现出来的亚符号表征上,并通过神经网络对数据表征的自学习和自组织产生恰当的回应性行为。不少研究者认为:深度学习技术已经超越了西蒙和纽维尔时代在符号系统中进行思维模拟的阶段,提供

① 如Allen Newell & Herbert Simon, "Computer Science as Empirical enquiry: Symbols and Search", *Communication of the ACM*, 1976(3), pp.113-126; Zenon Pylyshyn, "Perception, Representation and the World", in *Computation, Cognition and Pylyshyn*, D.Dedrick & L.M.Trick Eds., MA: MIT Press。

② Allen Newell & Herbert Simon, "A.Computer Science as Empirical enquiry: Symbols and Search", *Communication of the ACM*, 1976(3), p.123.

了一个更加优越的关于大脑神经网络的抽象认知属性的结构模型[①]。人工智能正在朝着超越人类智能的方向发展。

的确,有了深度学习技术的支持,人工智能在这个复杂的世界中表现得像个"超人":它在金融预测、自然语言处理、机器翻译等方面的能力远超传统的计算机,甚至在机器视觉和图像识别、语音识别、信息检索处等方面超越了人类的识别能力。深度学习向我们展现了一个物质系统是如何高效地对数据进行编码和利用的。从这个角度说,深度学习的技术革命无疑是成功的。但就研究现状看,今天的人工智能更像是一个弥补人类智能进化缺陷的、复杂而精细的"专门系统"。它还存在着一些亟待解决的问题,如无法应对小数据环境、无法进行迁移学习、无法进行常识性推理以及无法拥有不可预见的反应能力等。举个例子说,如果我们要让一个智能体通过深度学习模型学会如何在北京安全地驾驶汽车,那么,我们需要为这个模型输入大量关于汽车和道路的情景数据集以及具有明确人工标注的动作(例如,前进、后退和停止等),然后再训练一个神经网络如何在不同的路况和环境中用不同的方式行驶。为了学会如何在不同的情况下避免与迎面而来的汽车相撞,这个模型可能需要先被撞成千上万次。并且如果要让这个智能体离开北京到上海行驶,它就不得不重新学习大部分的知识。因为对深度学习模型而言,如果要将多层次的人工神经网络从原来训练的领域和环境迁移到新的领域和环境的话,就意味着要对绝大多数的参数(即权重)进行重新设置。

不难发现,尽管 GPS 与深度学习技术用来解释智能本质的方法不同,但两者有着共同的形而上学核心——根据内部表征(符号的或者是亚符号的)来处理信息。它们都自觉或不自觉地默认了客观主义的哲学假设。越来越多的研究者们意识到:人类智能的内在机制及其体现出来的实时性、灵活性和强健性是基于明确表征的人工智能研究范式所不能完全解释的。[②]

三 身体、行动与渗透的大脑

正是出于对传统表征和客观主义假设的不满,一条既不同于客观主义也不同于主观主义的中间道路——具身性的认识路线悄然发展起来。具身性的认知论质疑近代认知论的表征主义特征,从学脉上说,具身性的观念直接出自胡塞尔等人所开创的现象学。一方面,研究者们对胡塞尔的交互"主体间性"、海德格尔的"在世存在"(being-in-the-world)以及梅洛-庞蒂的"知觉—行动结构"等思想成果进行了深度挖掘。另一方面,研究者们积极吸纳心理学、神经科学、生物学等科学研究中有关具身效应的经验研究成果,如,内德森(P. M. Niedenthal)证实了记忆中情绪信息的具身特征;加塞莱(V. Gallese)发现知觉—行动系统的身体体验对大脑中动作概念的形成具有决定性作用;史密斯(L. B. Smith)和西伦(E. Thelen)以瓦特的离心调速器为例,强调了动力系统理论的重要性等,促进了具身性思想的形成。

具身性的思想观念具有革命意义。它关注身体以及身体、大脑和周围世界之间复杂的相互渗透关系在人类智能形成和实现中的重要作用,并从"主客融合"的哲学判断出发强调:(1)智能不仅仅以抽象运算法则的形式存在,而需要身体的介入;(2)智能体和世界之间的关系不是单向的认知关系,而是深刻的存在关系;(3)智能行为不是中央控制系统进行逻辑推理的产物,而是耦合系统涌现的结果。这种具身的智能观与经典的智能观形成了这样一种基本的反差:一种作为渗透的具身行动控制器的心智形象以及一种作为孤立的理性推理机的心智形象。

但是,具身性的智能观是否完全不同于经典的智能观呢?其中的一个核心争论就是关于表征在智

---

① Artetxe, M. Unsupervised Neural Machine Translation [Z/OL]. http://xueshu.baidu.com/s?wd=paperuri%3A%28d96361d322c2838c3625409eebf5b364%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Farxiv.org%2Fpdf%2F1710.11041&ie=utf-8&sc_us=12382315736483201531.

② 参见 Shaun Gallagher, *How the Body Shapes the Mind*, Oxford: Oxford University Press, 2005; Andy Clark, *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford: Oxford University Press, 2016。

能中的作用问题。激进的具身性观念主张表征无法对大脑、身体和世界之间复杂的耦合(coupling)进行充分解释,因此我们应当用动力系统理论(DST)来替代表征。然而,这种激进的结论无法得到辩护,也达不到预期的目的,因为这种怀疑的结论建立在对"表征"和"计算"过于狭隘和严格的解读之上。温和的具身性观念并不排斥"表征"和"计算"的概念,而主张将身体的相关属性表征在构成认知的算法中。研究者们提出:应当探究多种表征的形式,"通过不同水平的混合分析……以及解释工具,将认知动力系统与关于表征的计算的观念共同建构起来"①,以关注人类在现实生活中所拥有的基本的、实时的反应性智能。

例如,克拉克以低等生物蟑螂为例,表明人类和动物赖以生存的智能类型往往依赖于那些具身的、非符号性的表征。②蟑螂因为拥有很强的逃生技巧,而很难被人类或其他动物捕捉。它们能够:

瞬间感知由侵犯者的运动产生的风力干扰;

准确区分不同成因产生的风:侵犯者带来的风还是正常的微风或气流;

和其他蟑螂交换信息;

在试图逃脱危险的时候,不是简单随意地逃跑,而会考虑自身所处的位置、障碍物(如,墙壁或转角等)的位置、光线以及风向。

克拉克将蟑螂所表现出来的逃生技巧看作是基本生物性智能的一种类型。蟑螂的反应能力比条件反射式的"刺激—反应"模型更具智能。这主要体现在:它们将身体的行动视为问题解决过程的一部分,将局部环境的线索视为有效表征的一个部分。它们的认知行动并不是普遍有效,而是具有明确的认知任务,并且其中并没有涉及大量复杂的表征性计算。那么,蟑螂到底是如何逃跑的呢?克拉克解释道:蟑螂运用身体的两根尾须来侦察前方的风。它们的每一根尾须上面都长有细毛,以感知风速和风向。如果风以 0.6 m/s 或更快的速度加速,蟑螂便开始准备逃跑:这就是生物区别正常的微风和侵略者来临的方法。并且,它们能够积极迅速地在感知到危险后作出反应:一只静止不动的蟑螂需要 48 毫秒,运动着的蟑螂只需要 14 毫秒,而作出转弯回应的时间大约是 20—30 毫秒。它们的转弯行动以基本的神经元环路为基础,这些神经元环路能够根据蟑螂当前位置信息以及当下环境的状况来自动调节弯度的大小。同时,蟑螂通过大量腹部中间神经元获得关于风的基本信息,但是最终的行动则是其他感知环境特征的神经元集群调节的结果。③

可以想象,如果要对蟑螂精妙的逃生技巧进行模拟,人工智能研究者们可能会将这个复杂问题的解决过程拆分为三个子问题域:1)如果你受到了攻击,那么不要直线逃往墙壁;2)如果你遇到了障碍物,那么试着绕过它;3)区分来自侵略者的风和正常的微风。但是,问题在于:目前人工智能所擅长的数据分析和学习,无法产生实际逃生行为所需要的快速反应能力。换言之,在现实生活中,蟑螂并不是通过语言形式的、类似于文本的信息对熟练的应对能力进行编码的。它们的逃生行为能够比在大脑中模拟的计算活动更便捷、更有效地完成某些认知和知觉任务,令它们在逃离危险的时候不顾一切地往前冲。

蟑螂所具有的常识性的和生物性的知识,是当前最先进的人工智能体仍然缺乏的。那么,人工智能还需要做什么?我们认为,一种比较有前途的进路是关注身体在现实生活中熟练的反应能力,将与身体相关的属性表征在构成认知的算法中。

四　默会表征与人工智能

在前一个部分中,我们探讨了温和的具身性观念如何在理论上是适恰的。接下来,我们将聚焦于丹尼特关于默会表征的讨论,表明温和的具身性观念如何在经验上也是可行的。

---

① Andy Clark, "The Dynamical Challenge", *Cognitive Science*, 1997(21), pp.461-481.

②③ Andy Clark, *Being There: Putting Brain, Body and World Together Again*, MIT Press, 1997, pp.4-7; pp.4-7.

在《心智表征的不同类型》一文中,丹尼特区分了三种不同类型的表征:明确表征、隐性表征和默会表征。[①]丹尼特认为,当代人工智能研究过度强调了明确表征在智能中的作用,而未能认识到默会表征对于人工智能研究的意义。确实,明确表征是我们最为熟悉的一种表征类型。举例说,"猴子"这个语词对现实世界中的猴子进行了明确表征;"自行车"这个语词对现实世界中的自行车进行了明确表征;同样,一张猴子骑自行车的照片对"猴子骑自行车"这一事实进行了明确表征。这里的语词"猴子"和"自行车"以及猴子骑自行车的照片都属于明确表征的范畴,因为它们都是根据某种解释系统而具有语义内容的物理对象。在丹尼特看来,明确表征能够物理地存储于一个系统之中、被系统所用,并获得解释。

而隐性表征则是逻辑地内含于明确表征的另一种表征类型。例如,一个系统明确表征"杰瑞是一只老鼠",并且"所有的老鼠都有一条细长的尾巴"。那么,这个系统就对"杰瑞有一条细长的尾巴"进行了隐性地表征。从这个意义上说,即使我们从未考察过杰瑞的尾巴,我们也能够对"杰瑞有一条细长的尾巴"这一命题进行隐性地表征。

丹尼特还讨论了第三种表征的类型——默会表征。以计算器为例,丹尼特问:"一个计算器是明确地、隐性地还是默会地对正确的数字计算进行了表征呢?"[②]一个计算器并不是通过一些明确的计算法则而得出正确答案的。具体地说,当一个计算器在计算"2+2"的结果的时候,它无需依赖"2+2=4"的规则而直接得出"4"的答案。丹尼特说:"尽管计算器是一种能够为明确问题提供明确答案的设备,但是除了作为输入和输出端显示出来的对问题和结果的表征,它的计算过程并没有依赖内部的明确表征。"[③]计算器系统被设计成按照计算规则来做计算,但是这些规则本身没有在系统中得到明确表征。在计算的过程中,系统并没有对"2+2=4"这条计算规则进行直接存储、操作和运用。并且计算器也没有对这些计算规则进行隐性表征。研究者们只是将计算器设计成符合运算规则来运行。由此,我们可以说一个计算器默会地表征了如何进行计算。

按照丹尼特的解释,明确表征是一个系统为了完成某项任务或解决某个问题而将它们作为表征来使用。假设,有某种计算器是需要查阅计算表来进行运算的,那么,系统使用这个计算表的方式决定了这个计算表就是明确表征。相反,尽管默会表征物理地出现于系统中,但是却没有被系统以表征的方式使用。默会表征令系统呈现出某种特定的倾向(disposition)。计算器系统特定的机制令计算器具有根据输入得出正确输出的倾向——计算规则引导了研究者们对计算器系统的设计,但是计算规则本身却没有被存储在系统中,也没有在计算器的运算过程中被调用。我们可以说,计算器"知道如何计算"(arithmetic know-how),因为它的系统默会地对如何计算进行了表征。

"知道如何去做"的知识包含了那些如何去完成一件事情的信息。例如,如果一个人说他知道如何做翻腾转体,那么,这个人至少具有一些关于如何做翻腾转体的信息。当然,这个人可能需要通过大量实践才能学会翻腾转体(实践是获得"如何去做"知识的重要途径)。因此,默会表征是一种令一个系统能够按照"知道如何去做"的知识来行动的结构。换言之,默会表征中承载了大量关于"知道如何去做"的信息。计算器能够做运算,是因为它的系统中包含了关于如何做计算的信息。同样地,因为一个人的"系统"中包含了如何做翻腾转体的信息,所以这个人知道如何做翻腾转体。

我们知道,激进的具身性观念主张我们只能对命题性知识(know-that)进行表征而无法对那种"知道如何去做"的实践性知识进行表征。丹尼特反对这种观点。尽管这两种知识类型在系统中发挥着不同的功能,但是它们能够(并且也需要)在智能系统中得到表征。一个智能系统能够明确地或隐含地表征命题性知识,默会地表征实践性的知识。回想前文中克拉克对蟑螂逃生技巧的讨论:克拉克并没有否认表征对于生物性认知的价值,而是要表明传统智能研究所强调的明确的表征类型无法对某些重要的认知现象作出解释。按照丹尼特的理解,我们可以说蟑螂的逃生机制中包含了大量的默会表征。也就

①②③ Danial Dennett, "Styles of Mental Representation", *Proceedings of the Aristotelian Society*, 1983(83), pp.213-226; p.221; p.221.

是说,蟑螂的身体对如何进行逃生的生存技巧进行了默会地表征。这些实践性知识中包含了大量的细节——如何感知风速、躲开障碍物、如何转身,以及大量关于周围环境的信息。但是,正如克拉克所说,正是蟑螂的生物性特征令它能够完成任务。当然,与丹尼特所描述的计算器案例不同的是,蟑螂的逃生本领并不是被研究者刻意设计出来的,但却同样包含着某种目的论——它们是物竞天择的结果。

五 结 语

人类智能的探究和实现问题既是一个哲学问题,也是一个科学问题。人工智能一开始出现的时候,是为了要弥补人类自身劳动能力的不足。当前最先进的人工智能研究以神经生物学的大脑为原型,将大脑想象成一个拥有大量清晰资料的逻辑推理装置,将智能的本质视为对大量明确表征的计算。

后期的现代哲学试图将认识论和形而上学的框架从一种对立的严格逻辑中释放出来,引入主体身体的概念,以反笛卡尔的方式重新理解身体与心智以及认知主体与外部世界之间的关系。这在海德格尔、萨特、梅洛-庞蒂的现象学中得到了集中体现。这些观念的变革导致了纯粹意识的退场。80 年代以来的具身性的智能观念强调智能不仅仅以抽象的运算法则的形式存在,还需要身体的介入:身体以及身体、大脑和周围世界之间复杂的相互渗透关系在人类智能形成和实现中起到了至关重要的作用。换言之,尽管人类的智能紧密地依赖于计算和表征的大脑,但从本质上说仍然是具身的。

但是,这种研究焦点的转换和解释框架的拓展并不意味着要抛弃“表征”和“计算”的概念。导致当前人工智能研究困境的根源并不是在于它采用了计算—表征的框架,而在于它还无法将有关身体的信息涵盖进对智能程序的描述中去。这种对具身性观念和表征性分析的合并,主要是对以“计算为核心”的智能研究范式作出调整。

未来人工智能的发展将在两个方面受到具身性观念的影响。一方面,在解释框架上,经典的“表征”和“计算”观念通过“默会表征”的新概念得到提升和拓展,使得明确的和默会的表征类型都能得到相应的重视,令我们对人类和动物的常识性、生物性智能获得更强的解释功能。另一方面,在实现水平上,通过将默会表征纳入系统结构,能够使智能体最大程度地利用环境的提示而简化复杂的表征内容和计算过程。当然,想要进一步推进默会表征在人工智能中实现机制的研究,还有许多问题需要厘清:例如,对于智能体的智能性行动而言,有哪些实践性知识是更为基本的?在将默会表征纳入系统结构的时候,有无顺序?默会表征与明确表征以及潜在表征之间是否可能相互转化?对这些问题的澄清对未来人工智能的发展来说至关重要,亟待进一步的深入研究。

(责任编辑 付长珍 王成峰)