

人工认知的语境建构与适应性表征解释

魏屹东

(山西大学 哲学社会学学院,山西 太原 030006)

摘要:认知科学的发展使自然认知得到充分研究,人工智能的发展使人工认知成为广泛关注的焦点。然而,这两种认知形式是何种关系,如何关联,如何进行推理,科学上存在着“探索黑箱”,哲学上存在着“解释鸿沟”。鉴于认知发生机制的复杂性和解释上的困难,这里将适应性表征作为认知发生的内在机制和解释框架,据此来探讨人工认知与自然认知的统一认知架构、认知推理的形式表征、概念与语境模型。研究试图表明,认知系统,无论是自然的,还是人工的,均是适应性表征系统,具有自适应、自复制、自组织的自主性和能动性;认知主体,无论是碳基生物,还是硅基装置,都是适应性实体,其行为都是适应性认知。因此,认知或智能是适应性和表征性的统一,适应性表征可合理地说明或解决认知或智能生成问题。这意味着,在适应性表征框架下,人工认知通过模拟自然认知的结构、功能和行为,生成像人类认知具有的通用智能,这可能是新一代人工智能发展的方向。具体而言,这里使用基于语境的适应性表征方法论来修正通用认知架构与认知推理模型,将人工主体分为逻辑主体、搜索主体、决策主体、学习主体和问题-解决主体,建构了一个形式化的语境模型,力图解释或解决科学和哲学普遍关注的智能生成问题,并探讨了其中蕴涵的本体论、认识论和方法论意义。

关键词:认知系统;认知架构;逻辑建构;语境模型;适应性表征;人工智能

中图分类号: B81; TP391

文献标识码: A

文章编号: 1008-7699(2023)01-0001-17

一个认知主体(cognitive agent),无论是自然的(动物和人类),还是人工的(人工智能和机器人),从适应性表征(adaptive representation)的视角看,^[1]它必须能够应对外部环境的变化(适应)而自主行动(表征)。这是作为认知主体的一个本质特征,用哲学术语讲,就是表现为主体性和能动性。人工主体(artificial agent)要适应性地表征,它必然要有一个结构,这个结构连同其所有组成成分在认知科学中被称为“认知架构”——一个既在整体的系统层次运行,使认知主体能够与其周围世界进行有效的交互,又在组件层次运行,展现出所有部分应如何组合以创建一个内聚的整体。^[2]⁸⁶根据安沙科夫(O. Anshakov)和盖尔盖利(T. Gergely)的说明,认知架构一般是由感知子系统、信息加工子系统、发生子系统和长期存储子系统构成的一个通用认知模型。^①^[3]⁴¹⁻⁴³接下来,笔者在“适应性表征”概念框架下,^[4-11]通过对已有通用认知架构及其符号表征的解析,对这些问题作进一步的探讨和分析,力图建立人工认知适应性表征的语境范式,旨在为解释或解决认知或智能的生成问题提供理论思路或哲学洞见。

一、认知系统作为适应性表征系统

认知,包括智能(自然的和人工的),是如何产生的,这是智能生成问题。该问题一直是哲学、心理学、

收稿日期: 2022-09-17

基金项目: 国家社会科学基金重大项目“人工认知对自然认知挑战的哲学研究”(21&·ZD061); 国家社会科学基金后期资助重点项目“人工智能: 从物理符号操作到适应性表征”(21FZX006)

作者简介: 魏屹东(1958—),男,山西永济人,山西大学科学技术哲学研究中心/哲学社会学学院教授、博士生导师,哲学博士。

① 在这个认知架构中,感知子系统负责接收来自主体环境的信息,如人的感官系统、人工感知器;信息加工子系统形成一个世界模型,主体使用这个模型作为决策过程的一个指南,如各种智能体、中央处理器;发生子系统实现主体的决策,并对环境产生影响,如促动器;长期存储子系统,即长时记忆系统,负责信息和数据的长期保存。

认知科学和人工智能面临的难题。认知涉及“感知”“意识”“自我”“心智”“智能”等心性概念,这些概念在认知主义范式中难以操作,在涌现论范式中则是模糊的。因为它们指称很难确定,或者根本就没有实存对象(仅仅是系统的属性)。

然而,一个不争的事实是,人是有意识和智能的,并能随着环境或目标的变化实时地调整其行为,笔者将这种能力称为“适应性表征能力”。那么,人工认知这种人造系统,也能像人那样随着环境或目标的改变及时调整和修正吗?能够像人那样具有适应性和灵活性吗?如何让人工智能也拥有类人的能力,是未来人工智能发展的关键。

要破解此难题,笔者认为要在概念框架上突破,这个概念框架就是“适应性表征”,一种主体运用中介客体对目标客体进行范畴化的能力,且这种能力随着目标与环境的改变会不断提升。如果这种思路和理解是对的,那么适应性表征既可作为认知或智能生成的内在机制,也可作为认知或智能生成的解释框架。这种假设不是纯粹的形而上学设想,而是基于如下科学事实。

第一,进化生物学表明,适应性是生物的一种普遍现象,自然选择或许还有文化选择一直在塑造着生物认知的本性。^[12]而由生物进化来的特定行为模式,如配偶选择,类似于计算机中的子程序,致力于增加适合度的特定功能。由此推知,生物的身体是适应性的,其基于身体的认知能力同样是适应性的,进而非生物的人工智能也应该是适应性的,这是人工智能如何通过适应性表征呈现不断变化的世界以及如何生成智能的内在逻辑。这意味着,适应性不仅是一种生物本能,也是一种认知能力,因此“智能也是一种理性”,^[13]可对系统的先天结构和后天行为进行修改。所以,我认为认知和智能的实质就是适应性表征。

第二,文化人类学表明,人的智能和知识是自然进化和文化进化共同塑造的,进化意味着适应,适应的同时彰显属性(表征)。如果说是自然进化(基因)造就了生物大脑,那么文化进化(模因)则进一步塑造了心智。^[14]当代认知科学的“具身认知”“嵌入认知”“生成认知”“延展认知”以及“情境认知”纲领,都说明了身体运动系统对于认知系统的不可或缺性,揭示出认知活动是适应性表征行为。

第三,人工智能的快速发展进一步表明,人工认知虽不是进化的产物,但它是人设计和制造的,其适应性自然地嵌入了人的适应性。物理学揭示,许多物理系统,如温度计、恒温器,具有适应性,遵循着系统演化规律;^[15]人工智能因模拟人的思维和行为,具有自搜索、自繁殖、自组织、自复制、自提升能力,在相当程度上是适应性的。^[16]例如,深度强化学习能够通过仿真平台,让智能体像人一样看电视、玩游戏,在虚拟世界中能够训练智能体在没有预编程的情形下,适应不断变化的外部环境并完成各种任务,从而通过自我学习来解决现实世界中的问题。^[17]这种自我学习能力就是适应性表征能力。

概言之,适应性表征可使人工系统产生智能行为。这是一个从自然认知到人工认知再到自然-人工认知融合的过程,一个肯定—否定—否定之否定的辩证发展过程,在哲学理念上实现了认识论和方法论的一次飞跃。

在笔者看来,这种将适应性表征作为人工认知或智能的解释框架的优势表现在如下几方面。

从概念看,“适应性表征”包括“适应性”“表征”两个子概念:“适应性”意味着自动调节和自我繁殖(复制),“表征”意味着自主表达和意义呈现,两个子概念组合而成的综合概念则蕴含了主体的反应性和意向性、具身性和情境性、自主性和语义性的统一。

从范式看,适应性反映了认知科学中涌现论范式的自组织性,表征体现了人工智能中认知主义范式的计算表征性,两种范式结合而成的“混合范式”是目前人工智能领域公认和普遍接受的,明显体现了两种范式优势的整合。因此,“适应性表征”概念充分展现了认知系统的共性。

从解决问题看,适应性表征就是要让无身的智能体也能适应环境变化,如智能体的搜索、表征和学习基本上都是适应性的,虽然这种适应性不完全是生物学意义上的,但适应环境变化是它们达到目标的共同特征。简单生物的适应只是对自然环境的本能适应,而人的适应不仅包括自然,也包括文化和社会。相比而言,智能体的适应不仅是对人为设置的内环境的适应,也包括对外部环境的适应。比如自动驾驶

汽车,它是通过灵敏的感受器和快速的处理器处理不断调整其行为的结果。

从主体性看,适应性表征作为解释框架,是通过主体得以实现的。根据人类主体表现的推理、发现、规划、学习和解决问题的能力,笔者将人工智能中的人工主体分为逻辑主体、搜索主体、决策主体、学习主体和问题-解决主体。之所以这样分类,是因为逻辑主体体现了推理能力,搜索主体体现了发现能力,决策主体体现了规划能力,学习主体体现了理解能力,问题-解决主体体现了完成复杂任务的创造力。这五种能力共同构成了人工智能的适应性表征能力,其中,逻辑是其符号表征的基础,搜索是其发现的本质,学习是其提升能力的方法,决策是彰显其主体性的标志,问题-解决是其认知的目标。

总之,如果人工智能拥有了适应性表征能力,一种理性推理能力,它就能够提升智能体的主体性,强化智能体的情境性和语境性。因为适应性表征是特定情境中的共变关系,即:表征与其对象是随着环境的变化而动态地变化的,智能的生成和知识的获得正是通过适应性表征进行的,其中有无意识介入并不是必须的。这正是人工智能得以发展的逻辑前提。

二、人工认知的适应性表征语境整合

如果认知适应性是一种理性能力,人工智能体作为异于人类的人工主体,其行动完全是遵循规则的理性行为。在确定的、可观察的、静态的或完全知晓的环境中,智能体所选择的行动的方法有多种,依据这些方法,它们能够发现其目标,这种仅仅按照某些理性规则寻找目标的过程就是“搜索”。尽管这种搜索是无意识的过程,却是适应性的。只是这种适应性不是基于生物本能的,而是基于理性规则的,完全按照规则(逻辑的、概率的)行事,这就排除了情感、意志等非理性因素。在这种意义上,因机器人没有情感、没有意识而否认其有认知能力、有智能的观点就有失偏颇了,毕竟“心灵”“意识”“意图”“意志”这些传统哲学或大众心理学的概念是难以按照规则操作的。人工智能将这些非理性概念排除在其研究领域之外,不仅是可理解的,事实上也是一种研究策略。

如果将人的适应性搜索看作是“完全适应性”,那么人工主体的适应性就是介于完全与不完全之间的一种中间状态,笔者将这种中间适应性称为“拟适应性”或“准适应性”,以区别于人的“意向适应性”。毕竟人的适应性是基于生物学的,而人工主体的适应性是基于物理学和计算机科学的,前者是碳基构造的(碳水化合物),后者是硅基构造的(物理硬件加软件)。在认知科学中,这个问题就是“硬件要紧不要紧”的问题,即意识特征是否与构成物质相关。哲学家塞尔坚持认为,意识是一种生物学现象,智能机不可能有意识,不可能有思维能力,它只不过是处理符号的物理装置,根本不理解所操作符号的意义。^[18]根据塞尔的看法,生物的适应性是基于有意识实体的,人工主体由于没有意识,当然不可能是适应性的。这意味着没有意识的东西不具有适应性能力,这个结论与智能体也能适应性地搜索和表征的事实不符。

依笔者看,意识和智能是两个层次的东西。有意识一定会有某种程度的智能,而有智能(符号处理能力)则不一定需要意识,如机器人。这里存在生物学和人工智能包括机器人学之间的一个解释鸿沟,类似于查尔莫斯关于意识的难问题方面的解释鸿沟。^[19]在适应性的意义上,只要一个主体能够在变化的环境中不断调整自己的行为,最终捕捉到要发现的目标,或者解决了所要解决的问题,就应该认为他/它是适应性的,在表达上也是适应性表征。^①

从认知系统的结构和活动机制看,人的认知系统(湿件)与人工智能系统(干件)完全不同,后者只是模拟前者的功能,就像鸟与飞机的飞行原理不同一样。这是不争的事实,也因此才有了认知科学、脑科学与计算机科学、人工智能等不同学科的存在。但是,仅就适应性特征来说,这两种认知系统(人和机器)都具有,只是程度上有差异,比如就灵活性、简单性来说,人的适应性比人工主体的要好许多。正是在这个

^① 这种判断与“图灵测试”类似,即:只要不能区分与我们对话的是人,还是机器(实际上是机器与人对话),那么我们就不得不承认机器也会思维(即操作符号、解决问题的认知活动)。

意义上,笔者将人工主体的非生物适应性称为“拟适应性”,以区别于人的生物适应性(本能+理性),^①因为这涉及意识问题。

人们普遍认为,人的意识具有意向性(初级意识)和反思性(自我意识)这两个特征,因而人才是真正有意识的物种。而其他动物,特别是灵长类的猿、大猩猩等,如果它们不知道它们是谁的话(也许朦胧地知道),它们就不具有完整或高级的意识,至多具有低级的意识(基于生命本能)。人工主体仅仅具有指向外部目标的属性,如搜索目标,但是不知道它们是谁。也就是说,它们仅仅知道如何做,但不知道它们知道如何做。这与“我会开车”和“我知道我会开车”之间的区别是一个道理,“我会开车”是如何做的问题,“我知道我会开车”是我意识到我会的问题。

这完全是两个不同层次的问题,前者是原初意向性(本能)问题,后者是自我意识(自指)问题。换句话说,人有命题态度,如“我知道”“我相信”,其他动物和智能机则没有。这涉及知识论中关于 know-how(知道如何做的方法问题)和 know-that(知道什么内容的命题态度问题)的区分问题。进一步说,人工主体是通过相关性认知的,而人不仅通过相关性,更能通过因果性认知。相关性认知是一种关联性认知,是一种影响和被影响的关系;因果性认知是一种产生性认知,是一种引起和被引起的关系。相比而言,影响关系较之产生关系要弱许多。从这个意义上,在笔者看来,人工智能若能产生因果性的认知或智能行为,其认知水平会提升一大截。笔者预计,因果性的介入可能是未来人工智能发展的一个方向,也是个难题。^②

一般来说,在人工主体开始搜索(行动)所需的结果前,它必须先识别目标,并形成一個明确定义的问题,因为有问题才能有目标和结果。问题由初始状态、一组行动、一个描述行动结果的转换模型、一个目标测试函数和一个路径代价函数构成。问题的语境由一个人为设置的空间状态来表征,从初始状态到目标状态穿过状态空间的路径就是要搜索的解决方案,搜索过程采取的策略是:搜索算法将状态和行动当作不可再分的原子状态。在人工智能中,一般的树搜索算法考虑所有可能路径去发现目标,而图表算法避免考虑多余路径。我们根据完备性、最佳性、时间和空间复杂性,就可判断出其优劣。这些搜索方法对于不确定的、动态的和完全不知晓的情形还无能为力,这是另一个更加复杂的搜索问题,但可通过设置新的语境来解决。由此看来,设置新的语境对于智能机器人的适应性表征是多么的重要!^③

有鉴于此,笔者将人工主体划分为五类,适应性表征是其共同属性,或者说,适应性表征是统摄不同人工主体的一个概念框架。这是一个通过五位一体交互产生适应性表征从而生成智能的认知模型,其中每个主体都有其各自的能力和具体语境,它们通过适应性表征,共同构成了人工主体的智能。

三、通用认知架构作为适应性表征系统

如果将认知架构视为认知系统,那它一定是适应性表征系统。在通用认知架构^{[3]44-49}中(见图1),感

^① 从哲学上看,适应性与意向性、相关性和因果性相关。意向性是关于或指涉某物的特性,最初是胡塞尔用于说明意识本质属性的概念,即有意识的物体就应该有意向性。但是这个概念用于说明意识太宽泛,因为一方面,简单生物(如阿米虫)虽然有生命,但没有意识;另一方面,即使是非生命的物体,如温度计、智能机,也具有意向性的特征,如指向某个目标,好像它们是有目的、有意识的。所以,在我看来,用意向性来描述意识的特性还不够,需要加上反身性或反思性,即知道“我是谁”(自我意识)。这样一来,意识就有了向外和向内两个特征,缺一个就不能准确说明人类意识的本质了。如果说意向性是某物有意识的必要条件,那么反身性就是某物有意识的充分条件。所以,对一个问题或概念的完备、准确的说明必须满足其必要性和充分性,这就是我们常说的“充要条件”。

^② 笔者预测,未来的人工智能还是离不开计算与认知关系的研究,因为认知科学和神经科学的大量实验表明,认知的基本单元不是计算的符号,也不是信息的比特,而是一种整体性的“组块”,如格式塔。所以,认知和计算的关系应该从四个方面进行研究:意识的认知机制——什么是意识?意识的神经表达——意识是如何在大脑中产生的?意识的进化——意识是如何起源于进化的?意识异常——精神疾病的本源是什么?参见陈霖:《新一代人工智能的核心基础科学问题:认知和计算的关系》,《中国科学院院刊》2018年第10期。

^③ 比如2022年11月30日诞生的人工智能聊天机器人软件 ChatGPT,据说有了可与人类匹敌的语境能力,即能够回答人类对话者提出的各种不同问题(不同问题体现了不同语境的存在),这被认为是深度学习的一个革命性成果。从设置语境的角度看,深度学习中的上层信息就是下层信息的语境,这样就可形成人类可以理解的语义。所以,让机器人拥有语境能力是其产生类人或超人智能的关键所在。

知(接收)子系统的功能是接收来自环境的信息,其中的物理信号和信息部分是明显区分的。因为接收来自物理环境的信号会形成数据(观察和实验的),而信息环境被理解为一组其他认知主体及其组合,其中一个主体可将信息转移到另一个主体(程序或算法),感知子系统可将这些信息转换为数据或知识。发生子系统执行一些行动,且这些行动被看作是对环境影响的反应,这是因为认知被认为是主体自主知识获得的过程,即主体独立实施其自适应可预测行动的能力。^{[2]3} 认知主体的任务就是获得知识,其行动也应该是促进这种认知行动的。长期存储子系统存储认知主体的离散活动的历史(过去经历的片段记忆),可被看作是一系列知识状态。

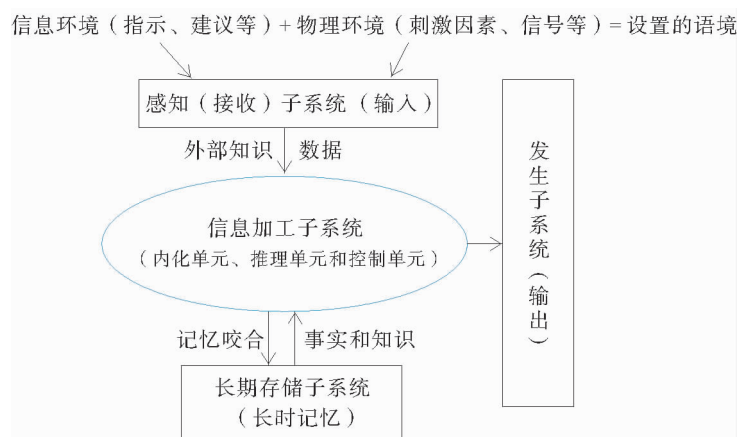


图1 认知系统的通用模型

注:根据 GERGELY T, ANSHAKOVO M. *Cognitive Reasoning: A Formal Approach* 结合语境模型绘制。

显然,信息加工子系统是这个认知架构的核心,其中的内化单元被用来积累来自环境的信息,并将其转换为认知主体的内在表征;感知子系统将以外在模板形式存在的“事实和知识”^①转化到内化单元,并集中到相应的缓冲区,如知识缓冲区和数据缓冲区。这种累积可以与认知主体的其他行动(如认知推理)平行地发生,必要时,来自缓冲区的知识和事实可以由相应的变换器(知识变换器和数据变换器)转化为认知主体的内在表征(其中数据和外部知识被称为事实,内在知识的这种转换就是数据和知识的内化)。

推理单元是信息加工子系统的核心,包括事实数据库(以下简称“事实库”)、知识数据库(以下简称“知识库”)和推理器(推理机制),事实库和知识库共同形成短时工作记忆。在事实库和知识库运行期间,推理器在工作记忆中引入了变化——知识库中的变化可被解释为新知识的获得,事实库中的变化可被解释为预测性假设的产生。对于推理机制的形式描述,知识和事实是以简单语句或陈述的形式被表征的,也就是我们常见的命题表征。

安沙科夫和盖尔盖利认为,事实和知识的表征与这种类语法表征的不同之处在于,一方面,事实可被表征为陈述,陈述是可直接被检验的。如果我们进行观察或实验的话,事实通常是观察或实验结果的表征或陈述,这是科学活动的常态。另一方面,在主观领域,知识表征了自然规律性。虽然主观领域(如思维过程)是不能直接被观察的,但可通过推理被发现。比如“该事件在这个事件后发生了”这个陈述是一个具体的事实表征,但“该事件是这个事件的原因”这个陈述表征了一个一般规律。这是因果规律的一般

^① 这里的“事实和知识”都是范畴化的对象,即使用语言描述的东西。按照安沙科夫和盖尔盖利的定义,“事实”是在一个系统起作用的语境或某种方法论应用的语境中以具体的基本陈述形式被表征的陈述,这种陈述没有参数。“知识”是这样一种陈述,其意义是在一个被考虑系统起作用的语境或某种方法论应用的语境中关于某些一般规律性的反映,而不考虑陈述的表征形式(是符号的,还是模型的)。这意味着,一个相同的陈述既可被看作一个事实,也可被认为是一种知识,如“地球上的物体自由下落”,这个陈述既是一个事实,也是关于这个事实的一种知识;既可以用自然语言描述,也可用符号表达,如自由落体公式。

表述,形式表达就是 $P \rightarrow Q$ (P 是原因, Q 是结果)。^①这个因果律我们不能直接看到,但它可以被揭示出来,如新型冠状病毒感染的发生可通过归纳推理得出。

这种表征事实和知识的陈述蕴涵了真值(事实的真假),这种真值在推理器的操作过程中是变化的,且这种变化随着确定性的增加是定向或有针对性的。这意味着,推理器的操作过程发生在事实库和知识库的不确定性减少的范围内。在信息论中,不确定性的减少意味着有序程度的增加,因此,这种不确定性的减少可被看作是从未知到有知的变化过程。事实上,推理的目的就是获得确定知识而减少不确定性,比如贝叶斯方法就是为获得确定知识的概率处理方法。

当然,这种推理过程的表征是实际情境的理想化和简单化。在逻辑中,如果认知主体有权处置其表征了事实和知识的所有可能的原子陈述,那么推理过程可以被还原为真值的变化。然而,在实际问题的推理中,一个主观领域的最终片段数是非常大的,穷尽搜索并处理它们几乎是不可能的。因此,在推理过程中,表征事实和知识的原子陈述得以形成,这就是认知主义的实证主义方法论。^[20]

在人工智能中,推理器的操作是由一组规则组成的算法执行的。算法也是一种知识,只不过是一种元知识(基于逻辑规则),这种元知识(如基本假设)可重复使用,但不能随意改变。算法的操作产生相同类型的重复程序的执行,由于这种重复,算法本身不变,但当推理单元返回而操作完成时,推理器完成了一个循环后,算法可被改变。根据这种操作,知识库的结构不是同类或同质的,而是异类或非同质的。一方面,由于推理,有一部分知识库可被改变,这一被改变的部分包括由原子陈述形式表征的专门知识,而原子陈述是关于某些事件、现象、结构和功能特性等之间的自然规则性;另一方面,在推理器执行一个操作循环中,有一部分知识库保持不变。变化的部分包括一些可从已接受且不会被怀疑的专业领域产生的先验知识,如相对论的光速不变假设。这些假设是作为元知识存在的,如逻辑中的矛盾律、同一律,它们是推理器进行推理操作的基础。

根据这个认知模型,事实库也是异类的。因为推理器的操作循环中包含一个不依赖变化的部分,而且这个部分存在一些明显的事实,在某个主观领域与客体的结构描述相关联,如关于基因结构的描述。推理器拥有自己的控制机制来管理推理过程,但这个控制机制不同于控制单元,后者用于控制作为主体的认知操作。推理的控制机制包括推理算法的一个编译器,并按照推理算法要求的次序执行规则。控制单元用于控制认知主体的活动,也就是控制认知主体的操作模式的变换(闭与开)。事实上,这个认知模型区分了认知主体的两种操作模式:感知内化模式和认知推理模式。内化阶段是推理单元的更新,且这种更新能够影响这个单元的所有组成成分,以便内化行动将数据和知识转化为认知主体的内在表征。

从感知-行动视角看,发生在推理阶段的过程包括三个行动:(1)固有推理,即上述描述的推理机制的操作;(2)数据和知识缓冲器中的数据和知识的累积;(3)子系统的操作行动,其目标是接收来自环境的新数据和知识。这三个行动可平行或同时发生,在它们之间还有两个行动发生:一是在内化阶段完成后和推理阶段开始前,短时记忆的内容存储于长时记忆中;二是在推理阶段完成后和内化阶段开始前,长时记忆中的工作记忆内容的存储也类似地发生了。因此,对每个推理的操作循环来说,在长时记忆中,这些行动保持工作记忆的两个简单印象,固定两个知识状态:推理开始前的状态和推理完成后的状态,其中两个检验器被用于控制信息加工子系统的操作模式的转换,起始条件检验器用于检测推理器的启动必须要满足的条件(检测行动发生于感知的内化阶段)。

从适应性表征来看,这些条件的实现,意味着认知主体有足够的事实开始从它们获得知识,并通过定量方式来表达。如果这些起始条件得到满足,一系列行动就会得到执行。比如:起始条件检验器发送指令“锁住”数据和知识缓冲器,感知的内化过程停止(如数据和知识转化为认知主体的内在表征),添加事实和知识库的过程停止,工作记忆的简单印象被保留在长时记忆中,起始条件发送“启动”指令给推理器,

^① 这里的“ \rightarrow ”是“引起”“产生”的意思,不是“蕴涵”(如果—那么),也不是“继承”,所以区别于逻辑上的蕴涵和继承关系。

推理器开始工作。记忆是行动系列中的一个重要环节,因为它保存过去的信息并预测未来,是能够让认知主体为行动作准备的机制。^[21]

四、认知推理的适应性表征机制

上述表明,认知推理是基于概念或范畴的。而基于概念的认知推理(信息加工活动),一般来说,是由相互嵌套的内循环和外循环构成的。^①在笔者看来,正是这两个循环构成了认知推理的适应性表征机制。这种能够处理相关认知推理的信息加工循环模型被称为认知推理框架(cognitive reasoning framework, CRF),类似于生物系统自动调节的内稳态(homeostasis),能够利用反馈维护自身所需要的系统变量恒定。^[22]

对于认知表征而言,在安沙科夫和盖尔盖利看来,使用CRF可以支持两类理论:一个是开放认知或准-公理理论,另一个是修正理论。“开放”是指认知主体与其环境之间是相互作用的,可获得用于分析的新事实,采用并内化来自相似理论的知识,修正推理规则、语言和数据,以及知识的内在表征形式,这意味着开放认知理论将模型化认知主体操作的外循环。“修正”是指工作记忆的明显区分部分的移动或转换,这里的“转换”意味着对事实的分析和知识的获得,以及事实库和知识库的转换。

在模型化认知推理的过程中,新陈述不仅会得到增加,而且过程本身也会得到修正。因此,修正理论不仅表征了认知主体操作的内循环(这个循环与推理连接),而且使用模型化这个推理过程的特殊推理技巧,如修正规则。修正规则将修改嵌入推论中陈述的真值和推论本身,或者说,修正规则不仅可以改变推论,也可以改变一些陈述的真值。然而,这些修正规则并不是什么新规则,而是逻辑中对应的似真推理规则,如归纳推理和统计推理。这样一来,修正理论允许认知过程的动力学表征,可被理解为是从未知(不确定性)到知识(确定性)的运动。^{[3]53}

显然,认知开放理论有利于建构环境的模型,修正理论有利于认知过程中的理论范式的变化。事实上,认知系统与环境的相互作用是不言而喻的,否则就成为孤立系统了。认知过程的修正意味着变化,包括认知规则和认知架构的改变。因此,相比而言,修正理论更为重要。^②修正规则可以被组织为类似于一个推理算法的系统,该系统控制一个修正推论的建构。而推理分为不同阶段,每个阶段对应于事实库和知识库移动的一次迭代。修正推论的非单调性与进入推理的一些公式的钝化(不活动)相关,即非单调性与推理的修正相关,而那些真值已经发生变化的陈述被认为是钝化的。那些依赖于钝化的陈述也是不活动的,而不活动的陈述不能被用于进一步推理的前提,也就不参与推理的进一步建构。

然而,这不意味着在使用修正规则时,也排除了推理过程中的矛盾或不一致的可能性。事实上,修正推理通常具有相当不一般的时态矛盾的特性,即:出现在推理过程中的矛盾由于进行了推理的修正,后来消失了。当然,矛盾还会出现。修正规则的价值就在于,能够及时地不断修正那些出现的矛盾,从而解决问题,这个修正过程事实上就是适应性表征过程。因此,修正理论旨在化解推理中出现的矛盾。这样一来,修正推理就成为情境的一个适应性表征,其过程对应于推理器操作的一个循环。

若要描述认知主体的全部认知活动,还需要开放认知理论,即需要从外部世界获得信息(事实和知识)。这意味着,开放认知理论应该能够反映与认知主体活动相关的不同认知过程,也就是拥有修正理论的所有可能性,如推理器在不同阶段进行推理的表征方式,而且也能够反映对接收来自环境的各种信息从内部进行重新组织的可能性。在这个意义上,对开放认知理论作出某些限定是必要的,这是因为,作为

^① 内循环由迭代推理算法组成,重复循环直到目标达成;外循环表征感知的复制和推理阶段,从认知主体活动的开始到结束,一直在运作。这种嵌套循环作为一种理想模型,是主体的认知系统(内循环)与其环境目标(外循环)的相互作用模式,构成了认知主体的认知推理过程。

^② 修正理论包括两方面:一是关于事实库和知识库的公理;二是关于推论建构的规则。公理是事实和知识的类比,因此,认知主体的推理单元内容包括事实库和知识库,可以由修正理论来表征。

表征世界的人工模型,其结果只是对真实世界某些方面的反映。

从语言表征来看,任何理论都可由一组文本来表征,而且这组文本可以延伸和扩展,构成更大的文本(语篇)。基于这种理由,开放认知理论被定义为认知主体活动的历史,从初始状态开始到达现在任何时刻的目标。因此,开放认知理论是语境论的一种新形态,因为“历史事件”是语境论的根隐喻,意指事件的变化和发展,而事件是主体参与的活动。^[23]就每个知识状态而言,建构一个对应的修正理论是可能的。在这个意义上,我们可将一个开放认知理论定义为一对修正理论的终端序列,以便表征认知主体从活动开始到现在时刻的活动历史。

然而,问题在于,作为认知过程历史描述的开放认知理论,如何与认知架构的组成部分相对应呢?我们知道,认知活动历史的简洁描述存在于主体的长时记忆中。在长时记忆中,事实库和知识库的内容的简单印象有两种情形:短时记忆更新后和推理过程开始前,以及稳定性产生后(推理过程结束后)。事实库的内容和部分知识库的内容可被解释为关于该领域状态的一个集合,可用某种形式语言来表征。知识库其余部分的内容可被解释为一个形式规则系统,对应于可靠的、似真的推理规则。这样一来,认知主体的短时记忆的内容可被理解为用于推理的公理和规则的集合。

上述通用认知模型表明,认知过程具有离散的特性。在认知推理中,两个阶段连续交替转换。在推理阶段,认知主体与环境联系,但不考虑其推理中来自环境的信号。在感知阶段,认知主体连续地解决两个任务:一是将事实和知识转化为内在形式(数据和知识的适配器负责解决这个任务);另一是从新旧事实和知识形成短时记忆的新内容^①(认知主体的大多智力成分和作为记忆控制器部分的更新装置负责解决这个任务)。这意味着认知主体有两种获得新事实和新知识的方式:一是保守的方式,即在一个专业领域使用推理方法形成关于不同实体之间关系的假设,这是科学认知通常的做法;二是激进的方式,也就是从环境(物理的和社会的)中直接吸收信息。也就是说,认知状态从信息成分中接受陈述性和程序性^②知识,而从环境接收的信息发生在认知主体的感知阶段,即通过感知器官接收来自环境的信息。这个阶段可能存在事实库和知识库的一个剧烈的更新过程,如推理算法可能改变,数据和知识模式转换的算法可能改变,内在数据和知识表征的模式(事实和知识的描述语言)可能改变。

众所周知,认知推理是基于感知的,之后才是推理阶段。在感知阶段,主体关于世界的初始模型开始形成,尽管感知过程不完全是独立的,形成的模式也不是完备的,至少部分初始知识库在认知主体的努力下应该形成。在推理过程中,修正理论处于发展中,因为其公理集可以被修正,但建构推理的规则集不能被修正。也就是说,当推理器被创建时,修正理论的发展就开始了。这个过程贯穿于许多中间阶段,直至达到稳定状态时结束(即收敛到一个目标),作为推理结果的修正理论将保存于长时记忆中。推理阶段完成后,一个新的感知阶段就开始了。在这个阶段结束时,就会形成一个新的修正理论。接下来新的推理阶段又开始了,直到认知主体的认知活动结束为止。一般来说,认知推理是有限的,比如认知主体的长时记忆内容是修正理论的一个有限序列,这个有限序列就是“开放认知理论”。^{[3]58}或者说,开放认知理论是修正理论的一个有限序列,每个序列是认知主体在认知活动开始前和结束后直接获得的短时记忆的一个简单印象。

概言之,开放认知理论与认知主体的记忆内容等同,这个内容可以被认为是认知过程历史的简化描述,长时记忆仅包含这个历史的关键时刻。如果一个主体被限制于仅通过推理接收新信息,那么修正理

^① 在这里,短时记忆内容的更新是一个非常困难的任務,其解决包括三个行动:将新接收的事实和知识与库存中的事实和知识作比较,目的是检测同一性、相似性、差异性和不一致性;依据某个标准,从新旧事实库和知识库的联合中选择新事实和新知识;形成新的事实库和知识库,这些库的联合通常不总是被表征为某些语言的一致公式集,而且在任何情形下,事实库都应该是一致的。需要注意的是,数据和外在知识的内化只有在在一个新事实库和知识库形成后才能完成。只有在这种情形下,新数据和外在知识才会被嵌入认知主体的事实和知识的内在系统。

^② 比如,推理器的操作算法属于程序性知识,将外在数据和知识转化为内在表征的算法也属于程序性知识。

论对其而言,就是足够充分的。如果想反映主体对来自外部世界的信息的接收结果,那么我们可以使用开放认知理论。在开放认知理论中,认知主体的感知阶段对应于从一个到另一个修正理论的转换,而且这个阶段存在关于事实库和知识库的更激烈的更新。

五、认知推理框架的语境化修正

然而,认知主体如何控制感知阶段,CRF并没有特别说明,这是它的一个致命缺陷。在笔者看来,语境框架可以弥补这个缺陷,因为认知过程实质上是语境化的,认知主体的活动历史、认知系统外的物理和信息因素均可看作语境因素。从语境论来看,语境作为背景或基础是不可或缺的,只是在认知过程中不直接参与认知推理,但其间接影响的辐射作用是不能忽视的。在这里,语境化体现了一个从未知到有知的认知历史的表征过程。

根据CRF,认知推理过程就是一个从未知到有知的表征过程。这个过程使用了修正推理的特殊句法工具,修正推理的建构使用了修正规则,这使得不确定性领域在逐渐变窄变小,确定性在逐渐变宽变大。这些修正规则可以修正进入推理的陈述的真值,以及修正推理本身。^①上述表明,认知推理是动态过程,具有离散性和规范性。推理过程的表征结构可以被分解为几个阶段,每个阶段还可进一步再分解为有限数量的模块,每个模块可分为两个时期:快速时期(其中出现归纳和统计似真推理)和慢速时期(其中仅有可靠演绎推理)。

据此,表征动力学的结构有三个嵌入层次:阶段(stage)、模块(module)和时期(phase)。在表征的意义上,这些层次都离不开基于语言的陈述,即命题表达,这就需要谓词(predicate)的概念。谓词是携带语义的,通常是动词和表语形式。比如“昨天下雨了”,“下雨”就是谓词。而语义的确定离不开语境,如上例中的“昨天”限定了下雨发生的时间,这意味着,我们使用的语言有所指,包括真值判断(是否下雨)。

显然,这种经验陈述和通过似真推理获得的陈述(自然语句),在有效程度上,与由严格可靠演绎证明获得的陈述(定律)相比,存在一定的差异。前者依赖于认知主体的活动历史(语境),后者依赖于逻辑规则。历史的积淀和直觉公理之间的差别是显而易见的,这种差别的存在不是因为一些陈述比另一些确证得更好,而是因为陈述性质确证的类型不同。这意味着有两种不同的真值类型:内在的和外在的。按照安沙科夫和盖尔盖利的看法,内在真值用于经验陈述和以似真推理获得的陈述的评价,外在真值用于被认为是通过严格证明获得的陈述的评价。^{[3]61}根据这种划分,存在着经验-理论、似真-可靠的二元区分:内在类型对应于认知经验和似真成分,外在类型对应于理论和可靠成分。可以说,真值的内在类型倾向于直接作用于经验数据,如归纳推理直接从事实出发得出可检验的结论。至于外在类型,它是对已完成推理结果的评价,即事后证明,如对归纳结果的演绎证明,允许从外部世界推出关于经验的和似真陈述,这是一种元层次的认知推理。^②

安沙科夫和盖尔盖利进一步认为,内在真值至少有两种:一是不确定的或未知的,可用?表示,其真值是指对所探讨问题的信息一无所知;另一种是确定的或已知的,其真值处于未知与已知之间的某个程度上,用概率描述,可用!表示。这种划分与科学哲学中的分析与综合命题的区分类似,也与传统哲学中的“分析的”与“实际的”区分类似。实际真值可再分为事实的值和假设的值,事实在认知过程的开始就给定,假设是在认知过程中产生的。可以看出,外在真值倾向于在知识论的指称中使用,内在真值主要是在预测和评估观察、实验和似真推理的结果时使用,也可以说是在知识的经验层次上使用。

就一个具体的认知推理结构来说,每个模块处理一个内在谓词,如根据语言描述内在陈述所使用的

^① 需要说明的是,传统推理规则只可将陈述添加到一个推论中,但不能对业已存在于推论中的陈述起作用。因此,推论的传统概念不能通过使用传统推理规则被修正,如传统逻辑规则不能通过自己的规则来修正,或者说,公理本身不能由公理自己来修正。

^② 这种元层次的外在真值可分为经典逻辑的真与假,即T和F。我们描述和表征事实、知识的语言就是使用合取 \wedge 、析取 \vee 、蕴涵 \rightarrow 、等价 \leftrightarrow 、否定 \neg 。这些连接符是在外在真值集上定义的,或者说是蕴涵和表达真假的。

谓词。一个阶段包括一个结构的所有内在谓词的加工,加工谓词的次序由修正运算定义。在给定的修正运算中,所有推理的这种次序是相同的,也就是推理不会改变加工谓词的次序。^①当然,一个良好设计的修正规则系统应该总是允许完成修正阶段的推理任务,否则,这种系统就没有必要设计出来了。如果修正规则可用于最后一个以当下谓词符号表征的不确定陈述,那么一个模块的修正阶段就完成了。

接下来进入演绎阶段,在这个阶段,只有演绎规则可以使用。这种从不确定推理转换到确定推理阶段(即从似真推理到演绎推理)意味着某种稳定推理,甚至终止推理。显然,修正推论将可靠(演绎)和似真(归纳、统计)推理组合起来。然而,将这种组合置于同一个推理中并不是一个无聊的问题(我们通常会同时使用归纳和演绎)。一方面,这个问题可以通过上述两类真值来解决;另一方面,可以在推理中以显在形式工具自我评估那些陈述的有效度来解决。

相比于其他表达方式,在笔者看来,运算符 $?$ 和 $!$ 是更为简单的自我评估工具,因为这两个运算符可将内在真值的集合映射到外在真值的集合上。譬如,对于内在真值集合中的一个真值 α ,其形式表达为: $?\alpha = T, \alpha = ?, ?\alpha = F, \alpha \neq ?; !\alpha = T, \alpha = !, !\alpha = F, \alpha \neq !$ 。^②比如一个经验命题 φ ,用 $?$ 和 $!$ 表征就是 $?\varphi$ 和 $!\varphi$,其直觉意义是“未知 φ ”和“确定 φ ”。对于确定(知道)的情形,在知识论的语境中,这是一个有争议的话题,比如知道到何种程度?是真知道,还是假知道?因此,可能存在其他的表述,比如,“ φ 可能是似真的”,“ φ 在经验上是真的”,“ φ 是经验适当的”。这就是接下来要讨论的 CRF 的形式表征。

六、认知推理框架的形式表征

为了进一步修正 CRF,这里需要详细说明安沙科夫和盖尔盖利提出的一个思想实验及其谓词逻辑的表征。^{[3]62-68}他们假设了一个叫米可的机器人,被发送到一个遥远的星球上去探测、观察和作实验研究。^③由于米可的操作条件非常严酷,时间间隔对于它而言,是非常重要的。不过,它可使用航行日记表来记录所发生现象的具体时间、现象的类和真值,由于这些是已知的,其真值是 $!$,没有观察到的现象由于是未知的,其真值是 $?$ 。

从谓词逻辑看,构成这些现象的表可看作是内在谓词表征,是从一个给定专业领域的一组客体到对应的一组内在真值的映射。比如,米可的事实库包含一个双谓词“……已发生……”,如“在 8 点 10 分发生了风暴”;一个三谓词(……发生后……间隔 10~30 分钟,又发生……),如“在 8 点 10 分风暴发生后,间隔 20 分钟,又发生了海潮”。没有发生的现象预计不久会发生,如“10 点后可能有雷电”。米可的知识库包含似真推理规则的形式对应物,即修正规则。

为了形成这些规则,安沙科夫和盖尔盖利引入如下标记法: $A(d, e)$ 表示语句“在日期 d 发生事件 e ”; $F(d, e_1, e_2)$ 表示语句“在日期 d 事件 e_1 发生后,间隔 10~30 分钟,又发生了事件 e_2 ”; $C(e_1, e_2)$ 表示语句“ e_1 是 e_2 的现象学原因”。^④根据这种标记法,修正规则可表征为如下形式:

$$\frac{?C(e_1, e_2), \exists d_1 \exists d_2 (d_1 \neq d_2 \wedge !F(d_1, e_1, e_2) \wedge !F(d_2, e_1, e_2))}{!C(e_1, e_2)} \quad (1)$$

① 具体来说,当内在谓词的加工是按次序一个接一个发生时,一个阶段的紧接模块开始得到执行。当一个新模块开始时,推理就处于修正阶段,在这个阶段,我们可以使用修正规则和演绎规则,即可以将修正规则运用到不确定性的真值的每个陈述,如模块中的陈述。当使用一个规则不能改变真值时,重复使用修正规则是禁止的。也就是说,从模块的启动开始,就无需将任何一个修正规则用到每个不确定的陈述上。

② 我们当然可以使用别的表达方式,比如 $K_a\varphi$ 表示主体 a 知道 φ (K 表示知道, φ 表示命题或陈述), $\neg K_a\varphi$ 表示主体 a 不知道 φ ,只是这种表达没有体现真值。

③ 这意味着,机器人在恶劣的环境中必须学会生存下来,它首先观察登陆的星球的自然现象,尽可能发现规律性。米可的事实库有一个航行日记装置,来记录所观察到的自然现象;航行日记装置中有许多表,包含关于那个星球的自然现象的信息,而且在 10~30 分钟的间隔期,自然信息会一个接一个地发生。事实上,这就是机器人的生存语境:它生活在一个躁动不安的海边,海岸上有几个洞穴可供它藏身。如果需要,它还可以爬到海边的岩石上,也即机器人可以花费 10~30 分钟到达最近的洞穴或爬上岩石。

④ 我认为这种非公理化的谓词逻辑标记法在描述和说明不确定性情境时非常有用,且简洁明了。因为它既没有过多和复杂的数学运算及逻辑推理,还蕴含了基于归纳推理的创造性,故而接下来笔者的基于语境的适应性表征修正也采用了这种标记法。

$$\frac{?F(d, e_1, e_2), !A(d, e_1) \wedge !C(e_1, e_2)}{!F(d, e_1, e_2)} \quad (2)$$

$$\frac{?A(d, e_2), !A(d, e_1) \wedge !C(e_1, e_2)}{!A(d, e_2)} \quad (3)$$

规则(1)可称为归纳规则,即:如果主体不知道事件 e_1 是否是事件 e_2 的现象学原因,如果存在不少于两个的日期 d_1 和 d_2 ,当主体知道间隔10~30分钟, e_2 紧接着 e_1 发生,那么关于“事件 e_1 是事件 e_2 的现象学原因”的陈述就是似真的。在这里,“不确定的”被修正为“似真的”,所以真值? \rightarrow 被! \rightarrow 替代。

规则(2)和(3)可称为因果预测规则。规则(2)的含义是:如果主体不知道在日期 d 间隔10~30分钟事件 e_1 发生后,又发生了事件 e_2 ;如果在日期 d 发生事件 e_1 ,而且“事件 e_1 是事件 e_2 的现象学原因”的陈述是似真的,那么“在日期 d 事件 e_1 发生后,间隔10~30分钟,又发生了事件 e_2 ”就是似真的。

规则(3)的含义是:如果主体不知道事件 e_2 在日期 d 应该发生,如果知道在日期 d 发生事件 e_1 ,如果“在日期 d 事件 e_1 是事件 e_2 的现象学原因”的陈述是似真的,那么“在日期 d 事件 e_2 应该发生”是似真的。在这里,“知道”和“似真的”对应于相同的内在真值。当然,这种表征不一定十分精确,但可以肯定,它能够使我们使用非常简单的逻辑规则来适应性地表征所发生的现象。

机器人的知识库包含一个描述内在谓词的表,即“……是……的现象学原因”,比如“黑云是风暴发生的现象学原因”,“类月大卫星(large moon)是潮汐发生的现象学原因”。这种知识库包含的推理算法可详细说明:

谓词 C(……是……的现象学原因)在任何可能条件下,试图以真值! \rightarrow 代替真值?(对应于归纳规则1);

谓词 F(在日期……紧随着事件……间隔10~30分钟,发生事件……)也在任何可能条件下,试图以真值! \rightarrow 代替真值?(对应于因果预测规则2);

谓词 A(在日期……有事件……发生)在任何可能条件下,试图以真值! \rightarrow 代替真值?(对应于因果预测规则3)。

通过执行这些推理算法,机器人会发现,“海洋上空黑云的出现是风暴发生的现象学原因”“类月大卫星是潮汐发生的现象学原因”等陈述,而且知道这些陈述是似真的,因而使用表中相对应行的真值来替代。接着,米可会将因果预测规则运用到表中对应的行来预测这些陈述,即谓词 F 和谓词 A,并会用相对应的真值替换。

如果这种表征是正确的,那么机器人似乎知道一个事件的发生会因果地引起另一个事件,或者说,它似乎掌握了因果规则。比如,它似乎知道,海洋上空黑云的出现会在10~30分钟,会有风暴发生。就机器人的操作程序而言,黑云已经出现,机器人在航行日记中记录下这一现象。因此,机器人应该走到最近的洞穴躲避,等待风暴过去。

进一步假设机器人的知识库包含一组先验知识,这有助于它适应环境条件。从语境论来看,这些先验知识就是语境,特别是这个知识库能够包含一些先验模式,根据先验模式,新知识被组织成一个系统。更重要的是,在先验知识中,可以发现一组行为算法,这些算法规定了机器人执行确定的行动,如果一些行动发生或作出预测的话。比如,如果风暴被预测到,那么机器人应该到最近的洞穴躲避;如果潮汐被预测到,它应该爬到岩石上。

由此可见,如果采用推理算法的修正推理方法,机器人的认知任务可以在修正理论中反映出来。算法由某一序列内在谓词决定,这个推论的一个模块对应于一个内在谓词的往返移动,其阶段对应于上述内在谓词的所有有限序列的往返移动。在机器人米可的例子中,整个推论与包含三个模块的阶段相匹配。在修正计算中,以表的形式描述一个认知推理是可行且有效的。^①从外部看,机器人似乎懂得归纳推

^① 这个表包括六个方面:推理行的序列数、表征陈述的公式、基础(那些公式为什么能被包括进推理中的理由)、阶段数、模块数和表的行数(使推论的给定行无效或被阻止)。

理,知道因果关系,这正是新一代人工智能要极力发展的能力——机器如何像人一样认知。^①

总之,在人工认知系统中,认知过程是基于事实库和知识库而发生的。当使用修正规则时,真值?被真值!所代替,这是修正规则的语义解释。从句法上看,当一个修正推论被建构时,这种真值的替换对应于将公式!φ 添加到已包含公式?φ 的推论上。如果公式!φ 被简单地添加到这种推论上,就会立刻产生一个逻辑矛盾,因为!φ 和?φ 不能同时为真。假如这种情形是可能的,则意味着上述公式应该同时接受两种真值!和?。因此,当添加!φ 时,必须阻止?φ,适应性表征不能出现矛盾,不一致就不是适应性表征了。

七、认知推理的适应性表征形式刻画

为了避免出现矛盾,要阻止演绎地依赖?φ 的所有公式。这意味着 CRF 及其推理规则是有限的,对通过迭代产生新知识的复杂情形还无能为力。上述表明,适应性是自组织系统的涌现特征,表征是认知系统的符号表达特征,两种特征的结合形成的“适应性表征”实质上就是认知科学中涌现范式和认知主义范式各自优势的混合(混合范式)。^{[2]81} 这样一来,完全可用适应性表征来简化 CRF 架构。

我们知道,表征是以区分为前提的,即 a 表征 b 意味着二者是不同的两个客体,可表示为(a, b)。若用符号 AR 表示适应性表征,则 a 表征 b 的适应性表征就是 AR(a, b)(其中, R 意为表征)。笔者进一步将人工智能的认知主体(CA)分为学习主体(XA)、搜索主体(SA)、决策主体(JA)、逻辑主体(LA)和问题-解决主体(WA),其适应性表征的形式表达分别是:AR_{XA}(a, b), AR_{SA}(a, b), AR_{JA}(a, b), AR_{LA}(a, b), AR_{WA}(a, b)。

由于所有学习、搜索、决策、推理和问题-解决都是认知过程,所以它们都是某种程度的适应性表征行为。这样一来,推理作为认知过程,自然就囊括在适应性表征框架内,不必出现上述的谓词 C、F、A。这就极大地简化了推理的过程。

这里仍使用上述标记法和形式框架,适应性表征的形式表达为:

对于一般认知主体的适应性表征过程:

$$\frac{?AR(a, b), \exists a \exists b(a \neq b) \wedge !R(a) \wedge !R(b)}{!AR(a, b)} \quad (4)$$

公式(4)的含义是:如果主体不知道 a 能否适应性地表征可观察或不可观察事实或现象 b,而如果主体知道在 a 和 b 存在且明确区分的情况下(现象的或因果的),主体可以使用表征工具 a 适应性地表征 b,最终获得可经验检验的结果(经验上适当)。

对于不同的认知主体,如学习主体,公式(4)变为:

$$\frac{?AR_{XA}(a, b), \exists a \exists b(a \neq b) \wedge !R_{XA}(a) \wedge !R_{XA}(b)}{!AR_{XA}(a, b)} \quad (5)②$$

这样一来,复杂的认知过程通过适应性表征,就可以得到简单而合理的说明,不必要的复杂表征可以剔除。因此,在表征问题上,适应性表征就像“奥卡姆剃刀”,可将多余的形式表达(比如谓词表达)去除。^③ 可以说,作为认知过程的表征不论用何种形式,凡是与表征对象不适当的,都必须被去除。这里笔者使用的是“适当”,而不是“准确”“精确”“可靠”。因为表征作为中介客体,只是一种认知工具,使用者据此可由已有观察数据和事实、实验结果,对预期进行的新认知(如实验结果)作出合理预测。

① 李德毅院士从进化论、遗传学和细胞学等人类认知的生物学基础出发,跨越物质科学、生命科学和社会科学,探讨了机器如何像人一样认知的“具身智能”的“机器生命观”,参见李德毅:《机器如何像人一样认知——机器的生命观》,《中国计算机学会通讯》2022 年第 10 期。

② 其他认知主体的适应性表征与公式(5)类似,即分别用 AR_{SA}(a, b), AR_{JA}(a, b), AR_{LA}(a, b), AR_{WA}(a, b) 替换公式(4)中的 AR(a, b),用 R_{SA}(a), R_{JA}(a), R_{LA}(a), R_{WA}(a) 替换 R(a),用 R_{SA}(b), R_{JA}(b), R_{LA}(b), R_{WA}(b) 替换 R(b)即可。

③ 这里提供的仅是一种理念,技术细节需要具体的算法或程序,而不是几个简单的谓词表达。而算法或程序是人工智能专家要做的工作,比如使用贝叶斯方法,通过获得新事实或证据,更新主体的信念状态,从而达到适应性表征的目标。

在科学中,实验的目的就是要查明或检验已有假设或理论的表征是否符合预期,从而查明预测是否合理,查明所使用的表征方式如模型、图形以及数学公式是否恰当。因此,表征只有合适不合适、恰当不恰当、适应不适应,没有真实不真实、正确不正确之说。比如原子模型,没有任何观察与原子的几何形状有关,即原子模型并非实际的原子中所直接观察到的东西(如果借助仪器可观察)。模型等表征形式只是我们思维的工具,有助于我们去认知,使我们能以可理解的方式,将所观察到的事实或现象综合起来,并对要寻求的新的事实或现象给出合理预测。正如物理学家薛定谔在论及科学模型时所指出的,“随着我们心灵的眼睛探入越来越小的距离和越来越短的时间,我们发现,自然的行为与我们周围可见、可触的物体的表现完全不同,以至于根据我们的宏观经验所构造的任何模型都不可能是‘真实的’”。^[24]

适应性表征作为认知的概念框架,在概念层次,可以作为统摄不同认知理论的统一范畴。也就是说,所有表征都应该适合其所表征的对象,否则,就不能称其为表征了,人工智能也不例外。^[25]作为认知主体的行动者使用各种形式的中介来描述客体对象,这个过程实质上是认知过程。在不同的专业领域,认知表征的方式会有所不同,但认知架构是统一的。^[26]即使是人工智能,其认知架构除了基本组成是物理的和生物的差异外,也与人类类似。也就是说,认知系统,无论是生物的人脑,还是物理的电脑,其结构一般都包括感知系统(信息输入)、信息加工系统、信息存储系统(记忆系统)和信息输出系统,其运作机制是一个控制-反馈过程:感知(输入)→信息加工(中央系统)+信息存储→信息输出→感知(输入)……这是一个无限可循环过程,只要信息输入和输出不停止。因此,适应性表征是一个动态过程,可以是一次完成的开放过程,也可以是多次完成的循环过程,采取哪一种,取决于所表征的目标的复杂性。越复杂的对象,其表征的循环次数就越多,反之就越少。

显而易见,在符号推理层次, $AR(a, b)$ 是基于区分的意义在特定语境中进行认知活动的。在这个定义中,“区分”表明了认知过程中的两个基本客体,“表征”指明了它们之间的联系,“适应性”意味着它们之间的适应度,“语境”规定了认知推理的范围和基础。公式(4)及其扩展表明,在目标明确和表征工具适当的基础上,认知主体通过适应性表征,可以从认知的未知状态到达已知状态,从而完成表征过程。在表征过程中,使用的工具不论是什么——自然语言、图像、物理模型、逻辑符号、数学方程式——只是表征的不同方式,在达到目的意义上,都是适应性的。在这个意义上,适应性也是一种基于经验适当性的理性。^①

在人工智能中,表征几乎完全是符号的,如计算机编程语言、算法和程序。在笔者看来,这种表征方式恰恰反映了人类思维的高度发展,是除人类外的其他动物所没有的。为了简化这种抽象表征方式,笔者引入“适应性表征”概念,而且可通过人工认知主体(学习的、逻辑的、搜索的、决策的、问题-解决的)来实现。原因在于,“主体”本身体现了自治(自主)性、能动性和智慧性等特性,而符号表征是只有人才具有的能力。正是人具有了符号表征能力,才创造了计算机、人工智能和智能机器人。

问题是,这些人造装置作为认知系统是否是适应性的,是否使用了表征?我认为这是毫无疑问的。人工智能中的主体不是人类主体意义上的,而是指一种算法或程序,即一种 agent,^②而非 subject,但它具有智能自治性(系统行为不受环境决定的程度,以及系统因此决定自身目标的程度)。^[27]比如贝叶斯表征,就是依据新证据调整信念状态的方法。按照笔者的理解,既然人的认知是适应性的,那么人造的智能也应该是适应性的,因为人会想方设法让其像人那样是适应性的,所以人工智能就是模拟人类智能的物理装置。如果不是这样,人工智能发展的意义就不大了,因为不具有适应性的人工智能难以实现人的目

① 在不同的表征方式上,除了自然语言(命题或陈述表达)外,其余表征方式都需要加以经验适当性说明,即使用自然语言加以解释,否则,人们就不知道表征的是什么。这意味着,任何信息只有通过适应性表征的解读或理解,才是有意义的。这有力地表明,自然语言是最基本的表征形式,也是普遍交流和理解所必须的。即使非常明显的图像表征,如肖像、风景画,也需要说明是谁的肖像、哪里的风景。符号和方程式表征更需要说明每个符号所表达的含义,以及整个表达式要表达的意义。在科学理论中,所有符号表征都有自然语言的说明,如化学中的元素符号和分子式都是约定的。事实上,符号表征使认知过程更简洁明了,并不会减弱适应性。

② agent 可以是人,也可以是除人之外的可以实施行动的具有自治能力的实体,如智能机器人。这意味着,agent 蕴涵了一种中介性或代理性,因此人工认知就是一种代理智能(行为体或智能体),是人设计并制造的假装智能。

标。原因很简单,环境是变幻莫测的,人工智能要想代替人做一些事情,就必须能够适应变化的环境,如机器人在月球上的探测活动。也就是说,适应性不仅仅是生物本能,也是一种人类理性行为。

八、认知系统作为适应性表征的哲学意蕴

适应性表征作为解释性的概念框架,相比于其发生机制的技术细节,其哲学意义更为突出。因此,这里有必要从本体论、认识论和方法论加以分析。

在本体论上,认知主体与其所要探索的世界相互作用,其行动与外部世界的模型一致。这意味着,认知主体的行动效果依赖于其关于境遇的模型的适当性,也就是模型对于所表征对象足够充分。这种适当性依赖于知识表征和反映相关世界的实际境遇的特性有多完备,而实际境遇的特性与所探索的世界类型或世界观相关,^①也即知识的客观性依赖于它如何适应性地表征所依据的世界观。这种适应性或适当性依赖于认知主体所持有的本体论,如实在论或反实在论,它决定主体如何表征其所面对的境遇或现象。

在适应性表征意义上,本体论的目标是表征系统,而表征必须在生成系统行为时发挥积极的因果作用,其中实体或客体与其描述之间有一一对应关系,即映射关系——每个实体或客体有一个唯一描述(范畴)或指定的表征,也就是外部客体在概念上的指称。在表征关系中,中介与其描述对象都是实体,若忽略实体和其唯一表征之间的区别,我们可以将它们都看作客体(一个中立概念),即中介客体和目标客体。这样的话,一个客体既可以被看作是在本体论上被唯一地表征的实体,也可以被看作是这个实体的表征本身(实体的替身)。换句话说,以客体代替实体更具客观性。当然,客体可以再分为子客体。比如,化学分子作为客体,可再分为组成元素及其关系。所以,化学表征是一种典型的符号表征,诸如分子式、化学反应方程式、分子的立体结构。总之,适应性表征的本体论要关注至少两个问题:要表达存在什么(客体),所谈论客体由什么构成;这些客体的最普遍属性和关系是什么。这就是认知主体需要的“本体论承诺”,它是认知主体的观察行动得以实现的基础。或者说,主体的观察行动要与由知识库表征的本体论定义相一致。

在认识论上,上述推理原则和方法不仅能够处理产生新的信息并促进知识形成的过程,也处理信息和知识的组织和表征。因此,知识在信息加工和认知推理中起着重要作用。这一过程必然涉及知识的起源、结构、发展和有效性问题,这就是认识论或知识论。这里的认知推理是基于计算机的人工认知,其主体不是自然人,而是一种使用算法的智能体。这意味着传统认识论是不适当的,比如科学哲学家波普尔的认识论——猜想与反驳,因为人工主体不会像人类那样去主观猜想,而且主体人的知识获得与发展是在社会和文化环境中进行的,这是人工主体所不具备的。

根据上述原则和方法,所需的认识论应该与处理信息和知识的方法论、知识发现的方法、检验知识有效性的方法和增加知识库的方法相关联。这种新的认识论能够与特殊类型的实验计算认识论和进化认识论相关,也就是结合了计算机科学和进化生物学的认识论。前者可以通过认知推理程序处理知识发现过程,如定律的计算机证明,以及由推理产生的知识增长;后者可以说明知识变化的过程。事实上,这种认知推理过程是从未知知识和问题的近似解决到产生新问题的转换过程——初始问题→试探性理论→排除错误→新问题,这就是波普尔的知识发展原则。

从控制论来看,推理和知识表征均是控制的过程,由谁控制?当然是主体了,人类或人工智能体。因

^① 科学领域一般有两类世界观:决定论和非决定论。根据决定论,世界的构成成分包括物体、实体或事件最一般的特性和关系,可以由决定论的形式描述来表征,如牛顿力学这种经典表征形式。在非决定论的情形中,如统计力学和量子力学,这种表征需要某些非决定论形式主义,如概率论、模糊理论以及可能世界理论。两种世界观混合的情形也是有的,比如,一种境遇既可能存在事件的随机变化,也可能存在事件间的决定论关系,从而影响世界的最终状态,如微观粒子行为的随机性和表征其行为的决定论的数学方程式。因此,对于适应性表征来说,认知主体知道外部世界的实际境遇的类型,对于以适当方式作用于环境的效果是非常重要的。这表明,适当性也依赖于认知主体形成的世界模型。可以预计,主体的世界模型越适当、越灵活,它反映的环境的效果就会越可靠,也就会更适应于其环境变化,这就是适应性表征。

此,从根本上说,所有类型的认识论都是有认知主体的认识论。人工智能虽然没有人类意义上的主体,但有类人的智能体作为主体。这就是为什么笔者将 agent 看作人工主体的原因所在,旨在表明它们具有类人的主体性。一旦这种人工主体拥有了人类的主体性,人工智能拥有人类的智能甚至智慧、意识和生命也不是不可能的。

这种新型的认识论能够将人类认知与人工认知相结合,这就是笔者提出并坚持主张的“适应性表征”,其特征在认识论上表现为如下特点。

第一,认知推理的建模是非亚里士多德式真理观(反映论),而是整合了符合论、融贯论、实用主义和语境实在论的真理观。符合论可用于说明基于经典逻辑和非经典多值逻辑的事实库;融贯论可用于根据似真推理规则对所作假设的评估;实用主义用于说明对所作假设的接受和认知推理的功效;语境实在论用于说明认知推理所获得的意义。^[28]所以,单一的真理观在这里没有市场。

第二,认知推理的真值一方面遵循传统哲学和逻辑的二分法,另一方面引入新的元素。根据认知主体的特性,所有真值分为两类:外在和内在的。外在真值有真假两类;内在真值至少有两类:不确定性和确定性的程度及性质,根据概率理论或统计理论确定。这类似于科学哲学中的分析和综合命题的分类,综合真值可分为事实真值(认知推理开始时给定的)和假设的真值(推理过程中产生的)。假设的真值是根据似真推理规则产生的,如归纳、统计、类比和溯因的结果。推理的可靠性和确定性以及真值的真实程度与初始给定的事实和论据相关,也与推理过程中的使用规则的合理性相关。

第三,适应性表征框架综合了认知推理框架中由开放认知理论表征的事实和知识。如上所述,开放认知理论是修正理论的一个有限序列,表征了认知主体的当下事实库和知识库,也以非逻辑公理部分刻画了所观察的外部世界。在笔者看来,开放认知理论是一种语境实在论,它将认知主体的认知活动历史纳入了整个认知过程(从推理开始到结束),体现了认知过程的连续性和知识库的开放性。这就是笔者所说的语境设置问题,即人为人工主体的推理设置必要的语境,包括历史事实和知识、各种已有规则和理论。^[29]这是认知推理所必需的认识论基础,比如机器学习的训练过程就是在添加语境因素。

第四,认知推理包括运用各种逻辑规则,如命题逻辑、溯因、类比等,也包括一些假设和公理,如上述认知推理框架中的公理和迭代,它们与传统逻辑的结合使得认知推理动力学成为可能。在数据分析和预测的基础上,我们可根据经验启示法形成假设,再根据逻辑规则进行认知推理。因此,基于语境的适应性表征提供了认知主体进行认知推理的范畴框架。

在方法论上,适应性表征框架为认知推理中遇到的假设形成、问题-解决、知识表征和意义展现提供了有效的基础。假设形成是基于历史事实和已有理论的,不是凭空的臆造,这就是历史语境的作用。问题-解决需要满足一些条件,如问题域中客体及其关系的描述要使用修正规则和相对应的计算程序,而且这种描述还要使用适当的语言来表征数据、事实和知识,如自然语言和形式语言的结合。这意味着,适当的表征语言对于一个成功的表征框架是非常重要的,如人工智能中使用命题逻辑,还是概率逻辑,结果是不同的。

在决定论或非决定论(随机的)语境中,适应性表征框架对发现因果性、分析和预测问题的境遇等是十分有用的。因果性是科学发现(认知)的一个重要方面,^①比如,要揭示一种新药的治疗效果(有效、无效或副作用)的原因,首先考虑的是该药的化学成分和分子式结构,其次是用法和用量。这种发现因果性的过程是一种认知推理过程,因为我们要找到其中的因果关系,是什么原因导致了该药的何种结果。这

^① 例如,发现一个客体的属性在其结构关系中的原因、一个系统的行为在系统架构中的原因、一个事件的表现其他事件关系中的原因。

种发现过程包括实验结果、事实、知识、生物化学和医学理论以及临床经验等。^①

总之,适应性表征对于从环境或实验数据和事实中发现和提炼新知识和新规律是很有用的。适应性是所有生物体的一种本能属性,认知适应性是所认知系统的一种本质属性,语境作为框架,是产生意义所必需的基底。适应性、表征与语境的结合,形成了一个基于语境的适应性表征框架,该框架有望在整合自然认知和人工认知、消解认知解释鸿沟,以及在人工智能拥有类人意识和智慧方面,发挥重要作用。

九、结语

上述探讨了认知系统的认知推理框架、适应性表征机制和语境框架,这是人工智能生成智能所必需的。事实库和知识库是认知推理的基础,也就是人设置的语境。在此基础上,人工认知系统(包括不同的子系统)使用各种假设和规则,如修正理论和开放认知理论,通过推理和表征形成知识,从而产生智能。这个过程无论多么复杂,无疑是一种基于语境的目标指涉的表征过程。人工认知主体——逻辑的、学习的、搜索的、决策的和问题-解决的,其共同特性是适应性表征,这是一种基于人类主体的自治性、能动性的反映能力,或者是由人类主体衍生出的一种认知能力。这种适应性能力加上语境,即基于语境的适应性表征,人工智能生成智能的机制就可能得到合理说明。

当然,让智能机器人这种人工认知系统拥有语境,并能够在其中适应性地表征,这是一种类人的认知推理过程。这种推理过程不仅理论上是可能的,而且技术上实现的可能性也很大,目前的人工智能研究正朝着这个目标发展。一句话,只要人工智能能够完全适应性地表征,它很可能会拥有类人的智能甚至意识或智慧,因为适应性表征也是一种生成模型,可使用误差不断更新符号表征。^[30]这种类人智能是否应该加以限制,这是哲学、伦理学和法学要进一步探讨的问题。

参考文献:

- [1]魏屹东.适应性表征:架构自然认知与人工认知的统一范畴[J].哲学研究,2019(9):114-124.
- [2]戴维·弗农.人工认知系统导论[M].周玉凤,魏淑遐,译.北京:北京大学出版社,2021.
- [3]GERGELY T, ANSHAKOV O M. Cognitive reasoning: A formal approach[M]. Berlin: Springer, 2010.
- [4]魏屹东.人工智能的情境觉知:一种“在线”适应性表征[J].上海师范大学学报(哲学社会科学版),2022(1):109-121.
- [5]魏屹东.人工智能对不确定性的适应性表征[J].上海师范大学学报(哲学社会科学版),2020(4):97-113.
- [6]魏屹东.人工智能的适应性知识表征与推理[J].上海师范大学学报(哲学社会科学版),2019(1):65-75.
- [7]魏屹东.人工智能的适应性知识表征[J].上海师范大学学报(哲学社会科学版),2018(1):28-39.
- [8]魏屹东.溯因推理与科学认知的适应性表征[J].南京社会科学,2020(7):34-43.
- [9]魏屹东.认知的适应性表征:机制、特征与功能[J].南京社会科学,2022(5):17-25.
- [10]魏屹东.论适应性表征及其方法论意义[J].山西大学学报(哲学社会科学版),2022(1):21-31.
- [11]魏屹东.适应性表征主义:一种新唯物主义[J].理论探索,2022(1):29-36.
- [12]FERREIRA C. Gene expression programming: A new adaptive algorithm for solving problems[J]. Complex systems, 2001 (2):87-129.
- [13]王培.智能论纲要[M].上海:上海科技教育出版社,2022:34-35.
- [14]HOLLAND J H. Adaptation in natural and artificial systems[M]. Cambridge, MA: MIT Press, 1992.
- [15]CHEESEMAN B L, GÜNTHER U, GONCIARZ K, et al. Adaptive particle representation of fluorescence microscopy images

^① 在我看来,疾病的诊断就是发现病因的过程,特别是面对复杂的或未知的疾病,诊断认知是复杂和漫长的,如新型冠状病毒感染的发现、诊断和治疗。即使有了发热、咳嗽等症状,也不一定就是感染了新型冠状病毒,也有可能是患了普通感冒或病毒性感冒。这就需要各种检测手段,如核酸检测等。关于问题境遇的预测,如新型冠状病毒感染治愈后的情况,如有无后遗症、有无再度复发的可能性、有无感染他人的可能性;这就要根据已有的各种数据、事实、知识和经验作出判断。从适应性表征来看,新型冠状病毒不断演化而导致的毒株变异和毒性弱化,正是其为了与宿主共存而不断适应性表征的结果。

- [J]. Nature communications, 2018, 9: 1-13.
- [16] SHADMEHR R, MUSSA-IVALDI F A. Adaptive representation of dynamics during learning of a motor task[J]. The journal of neuroscience, 1994(5): 3208-3224.
- [17] KIM T, JEONG M, KIM S, et al. Diversify and match: A domain adaptive representation learning paradigm for object detection [EB/OL]. [2022-08-15]. https://openaccess.thecvf.com/content_CVPR_2019/papers/Kim_Diversify_and_Match_A_Domain_Adaptive_Representation_Learning_Paradigm_for_CVPR_2019_paper.pdf.
- [18] SEARLE J R. Minds, brains and programs[J]. Behavioral and brain sciences, 1980(3): 417-457.
- [19] PAPINEAU D. What exactly is the explanatory gap?[J]. Philosophia, 2011, 39: 5-19.
- [20] FREEMAN W J, NÚÑEZ R. Restoring to cognition the forgotten primacy of action, intention and emotion[J]. Journal of consciousness studies, 1999(11-12): ix-xx.
- [21] WOOD R, BAXTER P, BELPAEME T. A review of long-term memory in natural and synthetic systems[J]. Adaptive Behavior, 2012(2): 81-103.
- [22] CANNON W B. Organization of physiological homeostasis[J]. Physiological review, 1929(3): 399-431.
- [23] PEPPER S C. World hypotheses: A study in evidence[M]. Berkeley, CA: University of California Press, 1942: 232.
- [24] 埃文尔·薛定谔. 自然与希腊人 科学与人文主义[M]. 张卜天, 译. 北京: 商务印书馆, 2015: 102.
- [25] 魏屹东. 适应性表征是人工智能发展的关键[J]. 学术前沿, 2019(11): 29-37.
- [26] NEWELL A. Unified theories of cognition[M]. Cambridge MA: Harvard University Press, 1990.
- [27] BODEN M A. Autonomy: What is it?[J]. Biosystems, 2008(2): 305-308.
- [28] HEXMOOR H, CASTELFRANCHI C, FALCONE R. Agent autonomy[M]. Dordrecht: Springer New York, 2003.
- [29] 魏屹东. 语境同一论: 科学表征问题的一种解答[J]. 中国社会科学, 2017(6): 42-59.
- [30] 渡边正峰. 大脑的意识, 机器的意识: 脑神经科学的挑战[M]. 岸本鹏子, 安婷婷, 胡实, 译. 北京: 北京大学出版社, 2021: 224.

Contextual Constructs and Adaptive Representational Interpretations of Artificial Cognition

WEI Yidong

(School of Philosophy and Sociology, Shanxi University, Taiyuan 030006, China)

Abstract: The development of cognitive science has led to the full study of natural cognition, and the development of artificial intelligence has made artificial cognition the focus of extensive attention. However, there is a “black-box of inquiry” in science and an “explanatory gap” in philosophy as to how these two forms of cognition are related and how they are reasoned about. In view of the complexity of the cognitive mechanism and the difficulty of explanation, adaptive representations are used here as an intrinsic mechanism and an explanatory framework for cognition to explore the unified cognitive architecture of artificial and natural cognition, formal representations of cognitive reasoning, and conceptual and contextual models. It is shown that cognitive systems, whether natural or artificial, are adaptive representational systems with autonomy and dynamism which are self-adapting, self-replicating and self-organizing; cognitive agents, whether carbon-based organisms or silicon-based devices, are adaptive entities whose behaviors are adaptive cognition. Thus, cognition or intelligence is a unification of adaptation and representation, and adaptive representations can reasonably account for or solve cognitive or intelligence-generating problems. This means that in the framework of adaptive representation, artificial cognition generates general intelligence like human cognition by simulating the structure, function, and behavior of natural cognition, which may be the direction of the development of next-generation artificial intelligence. Specifically, a context-based adaptive representation methodology is used to revise the general cognitive architecture and cognitive reasoning model; the artificial agents are divided into logical agent, searching agent, decision-making agent, learning agent, and problem-solving agent; a formal contextual model is constructed to explain or solve the intelligence generating problems in science and philosophy. Meanwhile, the ontological, epistemological, and methodological implications are explored.

Key words: artificial cognition; cognitive architecture; logic construction; contextual model; adaptive representation; artificial intelligence

(责任编辑: 江 雯)