



“范式变革”引领与“信息转换”担纲:机制主义通用人工智能的理论精髓

钟义信

引用本文:

钟义信. “范式变革”引领与“信息转换”担纲:机制主义通用人工智能的理论精髓[J]. *智能系统学报*, 2020, 15(3): 615–622.

ZHONG Yixin. Leading of paradigm shift and undertaking of information conversion: theoretical essence of mechanism-based general AI[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(3): 615–622.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202002019>

您可能感兴趣的其他文章

人工智能伦理体系:基础架构与关键问题

Ethical system of artificial intelligence: infrastructure and key issues

智能系统学报. 2019, 14(4): 605–610 <https://dx.doi.org/10.11992/tis.201906037>

集对分析在人工智能中的应用与进展

Application and development of set pair analysis in artificial intelligence: a survey

智能系统学报. 2019, 14(1): 28–43 <https://dx.doi.org/10.11992/tis.201803030>

机制主义人工智能理论——一种通用的人工智能理论

Mechanism-based artificial intelligence theory: a universal theory of artificial intelligence

智能系统学报. 2018, 13(1): 2–18 <https://dx.doi.org/10.11992/tis.201711032>

泛逻辑学理论——机制主义人工智能理论的逻辑基础

Universal logic theory: logical foundation of mechanism-based artificial intelligence theory

智能系统学报. 2018, 13(1): 19–36 <https://dx.doi.org/10.11992/tis.201711033>

因素空间理论——机制主义人工智能理论的数学基础

Factor space-mathematical basis of mechanism based artificial intelligence theory

智能系统学报. 2018, 13(1): 37–54 <https://dx.doi.org/10.11992/tis.201711034>

从人类智能到机器实现模型——粒计算理论与方法

From human intelligence to machine implementation model: theories and applications based on granular computing

智能系统学报. 2016, 11(6): 743–757 <https://dx.doi.org/10.11992/tis.201612014>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202002019

“范式变革”引领与“信息转换”担纲：机制主义通用人工智能的理论精髓

钟义信

(北京邮电大学人工智能学院, 北京 100876)

摘要: 世界人工智能研究至今一直沿用着物质科学的科学范式(科学观和方法论), 因此被分解为结构主义人工智能(人工神经网络)、功能主义人工智能(物理符号系统/专家系统)和行为主义人工智能(感知动作系统/智能机器人)3个各自为战互不相容的学派。虽然各个学派都获得了一些精彩的局部性专用性成果, 却没有通用性整体性的人工智能应用, 更无法形成通用的人工智能整体理论, 这成为人工智能研究与发展的最大痛点。目前, 通用性整体性的人工智能理论越来越成为社会的紧迫需求。为此, 本文依据作者四十多年研究的积累, 总结和提出了“机制主义通用人工智能理论”, 特别强调了“范式变革”和“信息转换”, 希望引起学界的研讨和批评。
关键词: 人工智能; 范式变革; 信息转换; 整体理论; 通用理论; 智能生成机制; 显因素; 隐因素
中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2020)03-0615-08

中文引用格式: 钟义信. “范式变革”引领与“信息转换”担纲: 机制主义通用人工智能的理论精髓 [J]. 智能系统学报, 2020, 15(3): 615-622.

英文引用格式: ZHONG Yixin. Leading of paradigm shift and undertaking of information conversion: theoretical essence of mechanism-based general AI[J]. CAAI transactions on intelligent systems, 2020, 15(3): 615-622.

Leading of paradigm shift and undertaking of information conversion: theoretical essence of mechanism-based general AI

ZHONG Yixin

(AI School, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Because of the employment of scientific paradigm (i.e., scientific outlook and methodology) in the field of materials science, the global research on artificial intelligence (AI) has been divided into three branches, namely, structuralism-based AI (e.g., artificial neural networks), functionalism-based AI (e.g., physical symbol systems and expert systems), and behaviorism-based AI (e.g., sensorimotor systems and intelligent robots), which are independent from each other and mutually unharmonious. A decent number of results from each of the three branches have been achieved; however, no progress has been made in the global theory of AI, let alone the universal theory of AI, and this has become the biggest pain point in the research and development of AI. Presently, the universal and global theories of AI have gradually become urgent social demands. Accordingly, the article titled “The theory of mechanistic general artificial intelligence” has been presented on the basis of the research experiences of the author during the past four decades; the work mainly focuses on “paradigm shift” and “information conversion.” The author hopes that the views and results presented in the article could draw discussions and criticisms from the readers of the AI academic circle.

Keywords: artificial intelligence (AI); paradigm shift; information conversion; global theory; universal theory; mechanism for intelligence growth; explicit factor; implicit factor

收稿日期: 2020-02-26.

基金项目: 国家社科基金项目(18ZDA027); 国家自然科学基金项目(60873001, 60575034).

通信作者: 钟义信. E-mail: zyx@bupt.edu.cn.

科学范式引发的思考。在理想的情况下, 为了对某个学科展开深入系统的科学研究, 首先需要对这个学科有一个整体性的理解, 能够在宏观

上说明这个学科“是什么”；在此基础上，还应当有“怎样研究”这个学科的宏观指南。前者就是关于这个学科的“科学观”，后者是关于研究这个学科的“方法论”。科学观和方法论两者是一个有机的整体：有什么样的科学观，就会有什么样的方法论。通常，这个整体就称为“科学范式”，也称为“研究范式”，或者简称为“范式”。

科学范式是科学研究活动的最高层次，是整个科学研究活动的指导和支配力量。不管人们是否已经意识到，一切科学研究活动都必然要接受某种科学范式的指导和支配，正如人类的活动无不受受到某种世界观和方法论的指导和支配一样。所以，科学范式（科学观和方法论）在科学研究活动中的地位 and 作用，就如同人类活动中的世界观和方法论的地位和作用一样。

众所周知，人工智能是一类开放的复杂信息系统，是信息科学技术的高端产物。因此，人工智能的研究应当遵循信息科学的科学范式。但是，20 世纪中叶人工智能问世的时候，世界上还不存在信息科学的科学范式，这是因为，信息科学本身也是在 20 世纪中叶才快速崛起，还没有来得及总结和提炼出自己的科学范式。在这种情况下，人工智能的研究就沿用了当时非常成熟也非常流行的传统物质科学的科学范式（其实信息科学技术本身的研究状况也是如此）。

传统物质科学的科学观认为：1) 研究对象是“客观”的物质，与主体因素无关；2) 研究对象是“确定性”的物质，可以分解，可以合成；3) 对物质的研究应当关注的是它们的结构与功能。与此相应，传统物质科学的方法论主张：1) 应当“排除主观因素”；2) 对复杂系统应当采用“分而治之”的方法；3) “纯形式化方法”就可以表示和分析物质的结构和功能。

正是沿用了传统物质科学的范式，人工智能的研究就一直践行着“排除主观”、“分而治之”和“纯粹形式化”的研究方法，并一直持续到今天。

传统物质科学的范式对于人工智能研究的影响主要有两个方面：一方面，迄今世界人工智能研究所取得的“局部进展”，得益于传统物质科学的范式在人工智能研究活动中的统领地位，要不是因为有“分而治之”和“纯形式化的表示和分析方式”，那时的人们就不知道应当怎样去研究这样高度复杂的信息系统；另一方面，人工智能研究至今都无法建立通用性整体性的人工智能理论，也受害于传统物质科学的范式在人工智能研究活动中的统领地位，要不是实行了“分而治之”与“纯形式化”，或许通用性整体性的人工智能理论早就

完成了。时至今日，研究者应当遵循怎样的科学范式来建立通用性整体性的人工智能理论？本文的研究，就从这里开始。

1 人工智能的历史经验与教训

20 世纪中叶，人工智能的研究从酝酿阶段进入到正式起步阶段。一个现实的问题摆在了研究者们面前：实实在在的人工智能系统研究应当从哪里入手？

那时，十分成功因而也十分流行的科学研究方法论是“分而治之”；它指出：在研究复杂系统的时候，可以把它分解为若干相对简单因而比较容易研究的子系统，当这些子系统研究成功之后，把它们的结果合成起来就还原成为了原系统。作为复杂信息系统的人工智能，人们认为它的原型主要是人类的大脑，它们当然是极为复杂的系统，因此符合“分而治之”的条件。

那么，如何对人脑进行“分而治之”的模拟研究？

当时学界已有共识，系统具有 3 个基本要素：

1) 结构是系统的第一要素，因为若是没有系统的结构，就没有系统本身；2) 功能是第二要素，有了一定的系统结构，就可形成一定的功能；3) 行为是系统的第三要素，有了一定的结构和功能，系统就会表现出一定的行为。

于是，人工智能系统的研究便从模拟人类大脑第一要素——大脑皮层结构开始。这便是以 1943 年 McCulloch-Pitts 发表的神经元数理逻辑模型为开端的人工神经网络研究^[1-5]。这种研究的标志性特征是“模拟大脑皮层神经网络的结构”，因此被称为“结构主义人工智能研究”（也有不少人把它叫作“连接主义”，但“连接”不是“结构”的唯一特征，“连接”没有刻画被连接的“节点”的性质，因此，“连接”只是“结构”的一个不完整的下位词，“连接主义”的称谓不如“结构主义”确切，还是应当称为“结构主义”）。可见，那时虽然没有使用“人工智能 (artificial intelligence, AI)”这个术语，但人工神经网络却是从人脑系统第一要素开始模拟人类智能的真正开端。这一点，通过近期“深层神经网络”的表现，使人们看得更加清楚。

由于大脑皮层神经网络极为复杂，模拟人脑结构的人工神经网络研究进展艰难。于是，一些思想敏锐的研究者便把注意力转向了模拟人脑系统的第二要素——系统功能。这便是以模拟人脑思维的逻辑功能为目标的第二种人工智能研究。

1956 年，McCarthy 和 Minsky 等在美国麻省 Dartmouth 举办了为期近两个月的模拟人类逻辑

思维功能的学术研讨会,诞生了以计算机为硬件平台、通过编制软件来模拟人脑思维功能、以通用问题求解为目标的物理符号系统,而且创造了“人工智能”这个术语来表征这种研究。由于这种研究的标志性特征是“功能模拟”,因此被称为“功能主义人工智能研究”^[6-12](虽然后来也有人把它叫作“符号主义”,但“符号”只是用来表示“功能”的一类记号,符号所表征的本质特征是“功能”,因此还是应当称为“功能主义”)。

随着研究的深入,物理符号系统研究遇到了越来越严重的“知识难题(知识的边界模糊、知识的获取困难、知识的表示能力有限,知识的推理能力不足等)”,于是便把“通用问题求解”的目标收缩为面向专门领域问题的“专家系统”。即使如此,专家系统的研究也仍然不能有效地摆脱“知识瓶颈”的困扰。

于是,又有了面向系统第三要素的研究:行为模拟。这就是 Brooks 等在 1990 年展示的模拟六脚虫爬行行为的智能机器人^[13-15],它能在不平坦的地面上行走而不会摔倒,引起学术界的高度关注。这就是“感知-动作系统”的研究。它的标志性特征是“行为模拟”,因此被称为“行为主义的人工智能研究”。智能机器人就是“感知动作系统”的一种具体类型^[13-15]。

今天,人工智能的研究风生水起,异彩纷呈,其实全部都源自上述“三大学派”。迄今,这三大学派各自也都分别取得了不少精彩成果:

1) 人工神经网络的研究在模式识别方面取得了不少成果,例如文字识别、语音识别、机器翻译、图像识别(包括人脸识别、指纹识别、医学图像识别等)、故障诊断、组合优化问题求解等,特别是深层神经网络显示了很强的分类与学习能力。

2) 专家系统的研究产生了一批机器定理证明系统,例如:战胜自己的设计者的跳棋系统,特别是战胜国际象棋人类世界冠军的 Deeper Blue 系统,战胜两位全美问题抢答冠军的 Watson 系统,以及战胜 61 位世界围棋顶尖高手的 AlphaGo 系统,等等。

3) 感知动作系统(也称为 Agent)的研究产生了一批各式各样的智能机器人,例如:小冰、小 i、小度等,其中,最引人注目的是获得了沙特阿拉伯联合王国国籍的 Sophia 系统,在一定领域的自然语言理解能力令人叹为观止。

至今,三大学派各自演绎出的人工智能应用系统已经不计其数。特别是 2016 年人工智能围

棋程序 AlphaGo 战胜韩国围棋顶尖高手李世石以来,世界各地的人工智能研究与应用风起云涌,形成全球性的人工智能发展大潮。

耐人寻味的事实是:按照“分而治之”的原则先后问世的人工智能三大学派,在整个历史上一一直各自为战,互不认可,甚至互不相容^[16-17],始终无法形成人工智能的整体理论,更不要说通用性的整体理论,甚至找不到建立通用性整体性人工智能理论的途径^[18-19],因此始终无法达到人工智能研究应当达到的成熟状态。

客观地说,在人工智能处于发展的初期阶段,研究的注意力主要还停留在“局部摸索”阶段的时候,“分而治之”方法论还没有显露出明显的负面作用;但是,随着人工智能的研究不断向前发展,注意力从仅仅关注人工智能的某些侧面、某些局部的研究,逐步发展到越来越关注人工智能的整体理论,甚至越来越关注人工智能的通用理论的时候,“分而治之”方法论便成为了人工智能研究的最大障碍。这是因为,按照“分而治之”方法论思想把人工智能整体“分解”为各个部分的时候,就不可避免地把各个部分之间复杂而看不见的信息联系割断了,而这些信息联系正是整体系统的生命线;生命线被割断了,整体理论的建立便不再可能。显然,只有从源头上摆脱“分而治之”方法论,人工智能的研究才能健康发展并逐步到达成熟状态。

此外,沿用传统物质科学的科学范式,人工智能的研究贯彻了“纯形式化”的原则。因此,它的各种概念(数据、信息、知识、智能)和各种方法(如各种变换和推理)也全面实行了“纯形式化”的原则。然而,人工智能的核心课题是“智能决策”,只有形式因素而没有内容因素和价值因素的决策怎么可能有足够的“智能水平”呢?由此造成了现有人工智能系统的理解力低下以及结论的不可解释。这种情况表明:人工智能的研究到了一个“不能不变”的历史关头。

总之,作为信息科学技术高端产物的人工智能研究,起步于 20 世纪中叶的工业时代并接受了传统科学的科学范式有其历史的必然。现在,信息时代已经发展起来,信息科学的科学观和方法论也已经锤炼成型,人工智能研究转型换轨、回归信息科学的科学范式的时刻已经到来!

2 信息科学的科学范式

那么,什么是符合人工智能研究性质的科学范式呢?

前已述及,人工智能是开放的复杂信息系统,因此应当遵循信息科学的科学范式。经过半个多世纪的思考、探索和体验,笔者终于完成了对于信息科学的科学范式的提炼。图 1 就是笔者总结的智能系统的简化模型,包括人类智能系统和人工智能系统^[20]。这个模型虽然如此简单,却是足够典型,而且可以清晰地揭示信息科学的科学范式的基本特征。

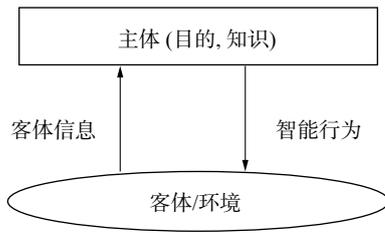


图 1 智能系统的最简模型

Fig. 1 The simplest model of an intelligent system

首先需要说明,环境中的任何客体都能产生信息,这就是它所呈现出来的自身的状态及其变化方式。但是,1)模型中所表现的“客体信息”并不是外部世界一切客体所产生的信息,而只是“能够直接作用于主体”的那些客体信息;2)并非所有能够直接作用于主体的客体信息都能参与“主体与客体之间的相互作用”,只有那些与主体的目的相关(正面相关或负面相关)的客体信息才能参与。换句话说,能够参与主客相互作用的客体信息是经过主体选择的,否则无穷无尽的客体信息就会使主体手足无措,被客体信息所淹没。所以,这种限定不但是必要的和合理的,而且完全符合现实的真实情况。

现在我们就可以来考察图 1 模型的工作过程了。

模型显示:人类主体是有“目的”的。人类共同体的永恒目的是要不断求得更好的“生存与发展”。同时,人类主体也具有先前积累起来的知识,通常称为“先验知识”。因此,当外界客体所产生的客体信息作用于主体的时候,为了维护主体“生存和发展”的目的,主体就必须生成相应的行为反作用于客体。这里对主体行为的基本要求是:它必须是智能的行为,否则主体就不可能实现自己的目的;同样,这个行为必须是智能的,否则就可能破坏环境的运行规律,而环境运行规律

的破坏也可能会反过来威胁主体“生存与发展”的目的。所以,主体生成的行为必需具有足够的“智能”水平。

这就是“智能”问题的由来:“智能”是人类实现“生存与发展”这个永恒目的的根本保证,没有智能,就没有人类的生存与发展。这就透彻地解释了:为什么“智能”问题会成为人类关注的焦点和热点。这是研究“智能”问题的全部意义所在。

而且,模型还表明,信息是智能(智能策略和智能行为)的真正源头。主体对信息的理解水平和把握程度如何,将在很大的程度上决定他所生成的行为是否真正具有智能以及智能水平如何。因此,忽视了信息作用的人工智能研究必定会“先天不足”。

传统物质科学的科学观的基本特征是“客观性、确定性”。但图 1 的模型却表明:1)研究对象应是“相互作用的双方是客体和主体”,而不是单纯的“客体”;2)相互作用的过程不是“确定性”的,而是充满了“不确定性”。这说明,传统物质科学的科学观确实不适用于智能系统的研究,智能系统的研究应当接受“主体客体互动性”和“过程不确定性”的科学观。

传统物质科学的方法论的基本特征是“分而治之,合成还原”。图 1 的模型却显示,在主体与客体相互作用的过程中,客体信息将被主体一步一步地加工生长成为智能行为,形成了一个以“信息生态,整体演化”为特征的信息生态演化过程。这说明,传统物质科学的方法论也不再适用于智能系统的研究,智能系统的研究应当转而接受“信息生态,整体演化”的方法论。

图 1 的模型还表明,在智能系统研究的场合,是主体驾驭着主体与客体的相互作用,主体关注的是它所追求的目的是否得到满足,后者必须采用“形式、内容、价值”的三位一体的描述和分析方法。可见,传统科学方法论的“纯形式化”方法不能支持智能系统的研究。

基于以上的讨论,我们就可以总结出信息科学的科学范式。

表 1 给出了传统物质科学和现代信息科学 2 种科学范式的基本特征。

表 1 两种科学范式的主要特征对比

Table 1 Comparison of main characteristics of two scientific paradigms

科学范式	传统物质科学的科学范式	现代信息科学的科学范式
科学观	客观性, 确定性	主客互动性, 非确定性
方法论	分而治之, 形式化	信息生态, 形式、内容、价值一体化

3 为什么要“范式革命引领”

本文开头便指出：科学范式处在整个科学研究活动的最顶层，是整个科学研究活动的最高指

导和支配力量。因此，在研究和建立一个学科的整体理论的时候，一定要遵循正确的科学范式。否则，就会事倍功半。表2给出了新科学研究活动中各种要素的合理层次结构^[21]。

表2 新科学研究活动要素的层次结构及其意义

Table 2 Hierarchical structure of the elements of discipline research activities and its significance

科学研究的要素		要素涵义	要素作用
科学范式	科学观	阐述一门学科的宏观总体认识	确保科学观念正确
	方法论	明确一门学科的宏观研究方法	确保科学方法有效
学科定位	全局模型	确立一门学科研究的基本架构	确保研究模型真实
	研究路径	确立一门学科研究的基本方略	确保研究路径合理
学科基础	交叉结构	认识本学科与其他学科的关系	确保交叉学科正确
	数理基础	明确一门学科理论的数理基础	确保数理基础精准
学科理论	基本概念	构建一门学科的基础概念体系	确保概念体系完整
	基本原理	构建一门学科的基本原理体系	确保概念联系通畅

表2的意思是说，一门学科的科学范式，是关于这门学科的“总纲”，它从宏观上界定了这门学科“是什么”（科学观）和“怎么研究”（方法论）。有了科学范式的引领，才能够明确这门学科的具体定位，包括明确学科需要研究内容的蓝图（全局模型）和研究的基本方略（研究路径）。有了正确的学科定位，才能进一步认识符合科学范式和学科定位所要求的学科基础，包括它与其他学科的相互关系和所需要的数学基础与逻辑基础。最后，根据所有上述要素的约束，才能确定学科所需要的基础概念和基本原理，形成学科的基本理论。

这是研究一门学科的有机层次结构体系。需要指出的是，人们直接看到的是由基础概念和基本原理所组成的“学科理论”，但是，在学科理论的后面却隐藏着一系列比它更基础、更深刻也更全面的要素：科学范式、学科定位和学科基础。而且，科学范式决定着学科定位，学科定位决定着学科基础，科学范式、学科定位、学科基础共同决定着学科理论。总之，在整个科学研究活动中，科学范式确实处于“统领”的地位。如果忽视了科学范式的指导和支配作用，只是单纯地考虑学科理论，忘记了根本，那就一定会出问题。特别是，在关注通用性和整体性人工智能理论研究的时候，遵循正确的科学范式乃是成功的根本保证。

就人工智能研究的历史与现状来看，人工智

能的范式处在“张冠李戴”的状态：传统物质科学的科学范式（张冠）支配着人工智能的研究活动（李戴）。这种状态对于通用性整体性人工智能理论研究显然极不合理。因此，颠覆传统物质科学的科学范式对人工智能研究的统领地位，确立信息科学的科学范式对人工智能研究的支配作用就成为首当其冲的任务。一旦摆正了科学范式的地位（正冠），理顺了上述要素的关系，整个科学研究就会一正百正，一顺百顺，一通百通。

4 为什么要“信息转换担纲”

在本文论述信息科学的科学范式的时候，借助了图1的“智能系统最简模型”。在那里，我们已经很明确地得到了“信息是智能的真正源头”的结论：一切智能策略和智能行为都源于外部客体所产生的客体信息的刺激，客体信息在主体目的的引领下、在知识的支持下，最终转换成了解决问题的智能策略和智能行为。没有客体信息的刺激，系统就不会产生对它的智能反应。

由此就可以从图1的模型提炼出一个规律：信息转换与智能创生。意思是说，首先要对客体信息进行一系列的转换（转换是必要的，因为信息本身毕竟不是智能的），然后才可能创生智能。这里提到了“智能创生”，表明生成智能并非易事，需要复杂的信息转换，需要创造性。

顺便指出，可以证明，在信息与智能的研究领域，“信息转换与智能创生”不是一个普通的规

律,而是一个普适性的基本规律,也就是“定律”,而且是与物质科学领域的“质量转换与物质不灭定律”和能量科学领域的“能量转换与能量守恒定律”相辅相成的三大定律。“物质不灭”和“能量守恒”是自然界的物质资源和能量资源领域的根本规律,必须严格遵守,不可逾越;与此相应,“智能创生”则是自然与社会的信息资源领域的根本规律,人们可以遵循和运用这个定律来生成智能,从而更好地实现“生存与发展”的目的。这 3 个基本定律恰好覆盖了物质领域、能量领域和信息领域,前两者告诉人们“不许如何如何”,后者告诉人们“应该如何如何”。综合起来,3 个基本定律告诉人们:在遵守“物质不灭”和“能量守恒”的前提下,可以通过“信息转换与智能创生”来为人类谋求更好的“生存与发展”前景。

现在,我们就可以进一步地追问:有了“信息”这个源头,怎样才能通过“信息转换”来生成智能系统所需要的“智能”(包括智能策略和智能行为)呢?

图 2 实际上是图 1“智能系统最简模型”的详细版:智能系统的标准模型,也是通用人工智能理论模型^[21-24]。

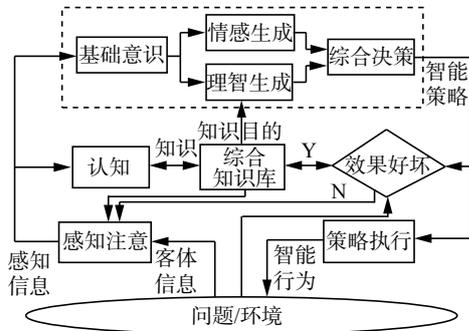


图 2 智能(人类智能和人工智能)系统的标准模型
Fig. 2 Standard model of intelligent (human intelligence and artificial intelligence) systems

图 2 和图 1 一样,都刻画了人类主体与客体相互作用的过程。图 2 中的椭圆表示的是环境及其中的客体(也可称为问题),其余部分则表示主体为了生成智能行为所必需经历的基本过程,也就是“信息转换与智能创生定律(即普适性智能生成机制)”的实施过程:

- 1) 客体(或问题)的信息作用于主体。
- 2) 客体信息到感知信息/语义信息的转换(感知与注意)。

由于主体具有目的和知识,所以,主体就通过“感知与注意”功能来判断这个客体信息是否与自己的目的有关,如果无关就不予理会(也就是“舍

弃”),如果有关(无论是正面相关还是负面相关)就予以关注并生成“感知信息”,后者具有 3 个分量:反映客体形态的“语法信息”,反映客体对于主体目的而言的利害关系的“语用信息”以及定义在语法信息和语用信息之上的“语义信息”。根据语义信息的这个定义,不难理解,语义信息可以代表语法信息和语用信息,因此也就可以代表整个感知信息^[21]。

3) 语义信息到知识的转换(认知)。

通过归纳运算,主体就可以把输入的语义信息抽象、提炼成为相应的知识。与语义信息的情形相应,知识也具有形态性知识、价值性知识、由形态性知识和价值性知识共同定义的内容性知识。这就是“通过认识活动获得知识”的认知过程。认知所生成的各种知识按照一定的规则组织起来就构成了包括本能知识、常识知识、经验知识和规范知识的综合的知识库。

4) 语义信息到智能策略的转换(谋略与决策)。

根据语义信息,主体就可以利用从综合知识库检索到的本能知识和常识知识生成相应的基础意识反应,利用本能知识、常识知识和经验知识生成情感反应,利用本能知识、常识知识、经验知识和规范知识生成理智反应,并将它们综合成为应对这个客体的“智能策略”(这个过程比较复杂,可以参见文献[23])。

5) 智能策略到智能行为的转换(执行)。

主体通过执行机构把智能策略转换成为“智能行为”,反作用于外部客体。

6) 偏差信息到优化策略的转换(检验与优化)。

在一般情况下,由于信息转换的每个步骤都存在不理想性和不确定性,把智能行为反作用于客体之后的结果往往与预期目标之间会存在误差。这种误差实际上就是一种新的信息,表示策略的智能程度还不够。因此,需要把这种误差信息反馈到智能系统的输入端,产生新的感知信息/语义信息,学习和提取新的知识,根据目的、新的信息和新的知识来改进策略,优化效果。这种反馈、学习、优化的过程通常要进行多次,直到误差足够小为止。这时所得到的策略才是真正可用的智能策略,把它存储在综合知识库(策略库)中备用。

7) 重设目标(更新)。

如果经过多次误差反馈、学习、优化过程都不能以满意的程度达到目标,主体就要考虑原来预设的目标可能不够合理,需要重新预设目标,重新进行上述信息转换过程,直到信息转换过程

所生成的智能策略能够满意地达到预设的目标为止。

以上7个步骤，构成了人工智能理论研究的基本问题，也反映了“人类认识世界和改造世界并在改造客观世界的过程中同时改造自己”的过程。人类就是在“不断地认识世界和改造世界并在改造客观世界的过程中同时改造自己”的过程中不断进步，不断求得越来越好的生存与发展。人工智能系统也就在这样的过程中逐渐实现主体目标。

从图2和上述7个步骤，可以非常清晰和确切地体会到：为什么在研究通用性和整体性人工智能理论的时候一定要以“信息转换”来担纲。实际上，主体生成智能策略和智能行为的整个过程就是实施信息转换的纲领的过程。

图2所示的智能系统模型是通用的模型，既适用于人类智能，也适用于生物智能和人工智能。当然，生物智能可以看作是人类智能的特殊情形，模型将要适当简化和特化，但原理是一致的。人工智能的情形也与此类似，特别是，人工智能系统的“目的”是人类主体赋予的。

通用模型的另一层意义是：它适合于解决各种不同领域的问题。它的通用程度主要取决于综合知识库所存储的知识领域范围：不管是什么具体的领域，解决问题的系统框架和工作机制是一致的，所不同的只是具体的信息内容、知识内容和智能内容。

在建立了通用性整体性人工智能的理论模型之后，接下来的问题就是要研究模型中每个模块的技术原理和系统实现。目前，这些问题已被解决。不过，由于篇幅的考虑，这些问题很难在一篇文章里都得到阐述。建议有这方面兴趣的读者可以参阅作者的相关论著^[22-24]。

5 结束语

处理好显要素与隐要素的关系。人工智能现今的各种应用，多少已经有点令人眼花缭乱，然而，对于人工智能下一步的发展来说，当前社会更为关注的却是通用性和整体性人工智能。而且，如果没有这方面的理论成果的进步，原有理论演绎出来的应用（如模式识别等）热情一旦减退，新一轮人工智能的寒冬迟早还会再度来临。

为此，本文主要根据笔者自己长期的研究积累，论述了通用性整体性人工智能理论研究的成功之路：“范式革命引领”和“信息转换担纲”。这是通用性和整体性人工智能理论研究最为关键的

两个问题：“引领”解决的是研究方向问题，“担纲”解决的是实施方略问题。

在表2所示的科学研究各种要素结构体系中，学科理论属于“显要素”，科学范式、学科定位和学科基础则属于“隐要素”。“显要素”是看得见的，“隐要素”是看不见的；但是，正如“思想指导行动”的道理一样，“显要素”是由“隐要素”决定的。这也是“有（有形的）生于无（无形的）”的道理。

所以，科学研究工作者应当同时关注显要素和隐要素，甚至应当更加关注隐要素，也就是要有刨根问底的精神。不无遗憾的是，人们往往只看见“显要素”，不太关心甚至完全忽视“隐要素”。这是缺乏创新特别是源头原始创新的思维根源。学科理论是怎么来的？学科理论应当怎样定位？应当遵守哪些重要的约束？比学科理论更为深刻的“科学范式、学科定位、学科基础”是怎样支配着学科的理论？只有正本清源，才能在源头上发现问题，在源头上创新，直至进入学术研究的“无人区”。因此，本文的讨论应当是有益的。

参考文献：

- [1] 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.
- [2] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. *Psychological review*, 1958, 65(6): 386-408.
- [3] HOPFIELD J J. Neural networks and physical systems with emergent collective computational abilities[J]. *Proceedings of the national academy of sciences of the United States of America*, 1982, 79(8): 2554-2558.
- [4] KOHONEN T. The self-organizing map[J]. *Proceedings of IEEE*, 1990, 78(9): 1464-1480.
- [5] MCCLELLAND J L, RUMELHART D E. Parallel distributed processing[M]. Cambridge: MIT Press, 1986.
- [6] MCCARTHY J, MINSKY M L, ROCHESTER N, et al. A proposal for the Dartmouth summer research project on artificial intelligence[R]. Dartmouth College, 1955.
- [7] NEWELL A. Physical symbol systems[J]. *Cognitive science*, 1980, 4(2): 135-183.
- [8] TURING A M. Can machine think?[M]/FEIGENBAUM E A, FELDMAN J. *Computers and Thought*. New York: McGraw-Hill, 1963.
- [9] NEWELL A, SIMON H A. GPS, a program that simulates

- human thought[M]//FEIGENBAUM E A, FELDMAN J. Computers and Thought. New York: McGraw-Hill, 1963: 279–293.
- [10] FEIGENBAUM A. The art of artificial intelligence: themes and case studies of knowledge engineering[C]// Proceedings of the 5th international joint conference on Artificial intelligence. Cambridge, USA, 1977: 1014–1029.
- [11] NILSSON N J. Principles of artificial intelligence[M]. Berlin: Springer-Verlag, 1982.
- [12] MINSKY M. The society of mind[M]. New York: Simon and Schuster, 1986.
- [13] BROOKS R A. Intelligence without representation[J]. Artificial intelligence, 1991, 47(1/2/3): 139–159.
- [14] BROOKS R A. Elephant cannot play chess[J]. Autonomous robot, 1990, 6: 3015.
- [15] BROOKS R A. Engineering approach to building complete, intelligent beings[C]//Proceedings of the SPIE Intelligent Robots and Computer Vision VII. Boston, USA, 1989: 618–625.
- [16] MINSKY M L, PAPER S. Perceptron[M]. Cambridge: MIT Press, 1969.
- [17] PAPER S. One AI or many?[J]. Daedalus, 1988, 117(1): 1–14.
- [18] RUSSELL S J, NORVIG P. Artificial intelligence: a modern approach[M]. Englewood Cliffs, N.J.: Prentice Hall, 1995.
- [19] NILSSON N J. Artificial intelligence: a new synthesis[M]. San Francisco: Morgan Kaufmann Publishers, 1998.
- [20] 钟义信. 人工智能: 概念·方法·机遇 [J]. 科学通报, 2017, 62(22): 2473–2479.
- ZHONG Yixin. Artificial intelligence: concept, approach and opportunity[J]. Chinese science bulletin, 2017, 62(22): 2473–2479.
- [21] 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 Yixin. Mechanism-based artificial intelligence theory: a universal theory of artificial intelligence[J]. CAAI transactions on intelligent systems, 2018, 13(1): 2–18.
- [22] 钟义信. 信息科学原理 [M]. 福州: 福建人民出版社, 1988.
- [23] 钟义信. 机器知行学原理 [M]. 北京: 科学出版社, 2007.
- [24] 钟义信. 高等人工智能原理 [M]. 北京: 科学出版社, 2014.

作者简介:



钟义信, 教授, 博士生导师, 发展中世界工程技术科学院院士, 国际信息研究会 2022—2023 年度候任主席, 北京邮电大学原副校长, 曾任国家 863 计划通信主题首届首席专家, 中国人工智能学会第 4 届和第 5 届理事会理事长, 国务院信息化领导小组办公室专家委员会常务理事, 中国科协—联合国信息技术咨询委员会主席, 主要研究方向为信息化、信息科学和人工智能基础。获得首届吴文俊人工智能科学技术成就奖和原邮电部科技进步一等奖等多项奖励。发表学术论文 500 余篇, 出版专著 18 部。