

机制主义: 人工智能的统一理论

钟义信

(北京邮电大学智能科学技术研究中心, 北京 100876)

摘要: 长期以来,学术界曾经分别从智力系统的结构、功能和行为三个不同的侧面分析和模拟人的智力过程,取得了众多成果,形成了人工智能理论的结构主义、功能主义和行为主义三大体系. 这些研究也留下许多问题,而且互相关缺乏沟通,常常发生孰优孰劣的争论. 本文试图提出智能生成的机制主义, 希望从智能生成的共性机制入手探讨智能的本质. 研究发现: 在一般情形下, 智能生成机制表现为信息到知识的转换, 由此引出了知识论; 特别有意义的是, 结构主义、功能主义、行为主义三者可以在机制主义框架下得到和谐完美的统一. 这可能为智能科学技术的发展提供新的理解, 带来新的机会.

关键词: 结构主义; 功能主义; 行为主义; 机制主义; 人工智能统一理论

中图分类号: TP18 **文献标识码:** A **文章编号:** 03722112 (2006) 022031705

Mechanism: a Unified Theory of AI

ZHONG Yixin

(Center of Intelligence Science and Technology Research, University of Posts and Telecommunications, Beijing 100876, China)

Abstract There have been so far three prestigious schools of AI theory respectively established through the structural, functional and behavioral approaches to human intelligence whereas there have been also, however, controversies among them over the theme of which one of the three should be the one who dominates. It is the author's belief that all the three schools did not cover the essence of intelligence. A new approach, mechanism approach, is thus proposed in the paper to cope with the situation. It is discovered that the mechanism of intelligence formation should be a series of transformations from information to knowledge and further to intelligence that are discussed in the paper. Interestingly but not surprisingly, the three existed schools of AI theory can well be unified under the framework of the mechanism approach. The mechanism approach and the unification of AI may hopefully provide a new foundation for the future development of AI theory and applications.

Key words structural approach; functional approach; behavioral approach; mechanism approach; knowledge theory; unified theory of AI

1 引言: 智能科学技术的意义

人,是一种典型的信息系统,也是典型的智能系统. 信息技术和智能技术的重要作用就是扩展人体信息系统和智能系统的功能. 如图 1 所示.

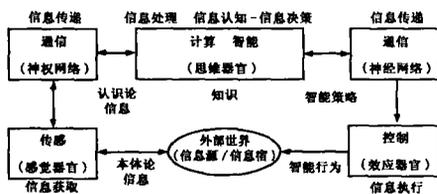


图 1 智能科学技术与信息科学技术的作用和关系

传感、通信、计算、控制等技术领域已经获得长足的进步. 但是,随着应用需求的不断深化,图 1 模型的核心——智能技术的发展提上了议事日程.

2 现有人工智能理论概述

现有的人工智能研究成果已经相当丰富多彩,从理论研究的角度看,这些成果主要由三个基本理论所驱动.

首先是由 1943 年 McCulloch 和 Pitts 发表的工作))) /神经元数理逻辑模型^[1]))) 所开创的人工神经网络方法. 它的基本思想是通过模拟大脑皮层神经网络的 /结构⁰特征来复现智能,因此,被后人称为 /结构主义⁰方法. McCulloch, Pitts, Rosenblatt, Widrow, Hopfield, Kohonen, Grossberg, Amari 等人是这一方法的代表性人物^[1~8].

收稿日期: 200520521Q 修回日期: 2005207206

基金项目: 国家自然科学基金 (No 6049632 No 60575034)

1956年一批年轻学者在美国 Dartmouth会议期间倡导了以计算机硬件平台支撑、用符号逻辑描写、由软件编程实现的/符号主义0方法。它不关心系统的结构特征,只关注系统的功能表现(因此也称为/功能主义0),回避了结构模拟的困难。由于功能主义方法一开始就用了/Artificial Intelligence(人工智能)0这一名称,后来许多人就把/人工智能0等同于/功能主义0或/符号主义0(其实人工神经网络所实现的智能也是人工智能)。McCarthy, Newell, Simon, Minsky, Feigenbaum等人是这一方法的领军人物^[9,24]。

20世纪90年代初期, Brooks等提出了/无需知识表示和推理的智能系统0的(行为主义0方法))在分析智能系统的输入(刺激模式)输出(动作模式)关系的基础上,系统首先鉴别输入的模式,然后根据输入输出之间的关系决定输出的动作方式。这便是模拟六脚虫的爬行机器人原理^[15,16]。

考虑到学术界对于人工智能理论中结构主义、功能主义、行为主义的具体研究成果和存在的问题早已人所共知,这里不再详述。

3 机制主义

虽然/结构、功能、行为0都是系统的重要属性,但是,对于智能系统来说,真正能够揭示系统本质的,却应当是系统的/工作机制0。于是,与结构主义、功能主义、行为主义方法不同,这里将直接关注/智能的生成机制0(为了简单,不妨称为/机制主义0)。

那么,/智能生成的共性核心机制0究竟是什么?

图1表明,/智能0的共性核心生机制可以理解为:在给定的问题环境目标的前提下获得相关的信息,并在此基础上完成由信息到知识的转换以及由知识到智能的转换(后者简记为/信息知识智能的转换0)^[17]。

我们之所以把上面的表述称为/智能生成的共性核心机制0,是因为任何智能的生成都会遵循/信息知识智能转换0这样的原则,只是/转换的具体过程0会随着问题的不同而有所不同;在某些相对简单的情况下,这个转换过程也可以有所简化。

4 智能生成的机制

既然智能的共性生成机制表现为/由信息到知识和由知识到智能的转换0,现在就逐一考察其中所包含的各种重要的转换问题。由于篇幅的原因,这里只讨论到/原理级0。

4.1 转换1:由本体论信息到认识论信息(信息获取)

智能生成机制首先要解决/在给定条件下获得相关信息0的问题,即本体论信息(外部世界的问题信息与环境信息)转换为认识论信息(系统获得的信息)的问题。为了研究这个转换,需要澄清/信息0的有关概念。

Shannon信息论^[18]认为,信息是消除不确定性的东西;不确定性是指状态和状态出现方式的不确定。例如对于具有N种可能状态 $\{x_n | n \in I(1, N)\}$ 的随机变量 X,若已知各个状态发生概率的分布为 $\{p_n | n \in I(1, N)\}$,那么相应的信息就可以用它的概率空间来描述:

$$\begin{bmatrix} x_1, \dots, x_n, \dots, x_N \\ p_1, \dots, p_n, \dots, p_N \end{bmatrix} \quad (1)$$

不过,这样的信息只能告诉/X的某个状态 x_n 会以概率 p_n 发生0,至于它与所关心的问题求解在多大程度上相关,却不得而知。因此,不能满足要求。

1988年,作者在5信息科学原理⁶^[19]中引入了本体论信息和认识论信息的概念。事物的/本体论信息0是指事物对其运动状态及其变化方式的自述;事物的/认识论信息0则是认识主体关于该事物运动状态及其变化方式(包括这些/状态/方式0的形式、含义和效用)的表述。其中,关于事物运动状态的形式的表述称为/语法信息0,关于事物运动状态的逻辑含义的表述称为/语义信息0,关于事物运动状态对主体所呈现的效用的表述称为/语用信息0。语法信息、语义信息、语用信息的综合体则称为/全信息0。可见,本体论信息是事物自身的信息,认识论信息才是主体所获得的信息。如果主体获得了事物的全信息,就不仅了解了它的形式,而且了解了它的内容和价值。

全信息需要三类参量来描述。对于一个具有N种可能状态 $\{x_n | n \in I(1, N)\}$ 的变量 X,用状态肯定度参量 $\{c_n | n \in I(1, N)\}$ 来描述 X 的语法信息(各个状态发生的肯定程度),用状态的逻辑真实度参量 $\{t_n | n \in I(1, N)\}$ 来描述 X 的语义信息(各个状态内容的逻辑真实程度),用状态的效用度参量 $\{u_n | n \in I(1, N)\}$ 来描述 X 的语用信息(各个状态对于用户的价值大小)。因此全信息的描述是与(1)不同的矩阵:

$$\begin{bmatrix} x_1, \dots, x_n, \dots, x_N \\ c_1, \dots, c_n, \dots, c_N \\ t_1, \dots, t_n, \dots, t_N \\ u_1, \dots, u_n, \dots, u_N \end{bmatrix} \quad (2)$$

其中 $0[()]_n [1, P, n, 2()]_n > < 1, ()$ 分别代表 c, t, u。符号 $> = <$ 表示/大于或等于或小于0。

既然事物的全信息提供了该事物的形式、内容、价值的信息,就可以判断这个事物的信息是否与当前所关心的问题相关,以及在多大程度上相关,进而就可以确定是否应当设法获取这个信息。文献[19]还给出了全信息量的测度公式。有需要者可以参考相关文献。为了简便,在下文的叙述中,只要不引起阅读的误解,我们把/全信息0一词简称为/信息0。换言之,下文出现的/信息0实际上都应当理解为/全信息0,除非另有特别的声明。

有了这些概念就可以讨论从本体论信息到认识论信息(全信息)的转换,图2示出了这个转换的基本原理。这实际上就是从外部世界获得相关信息的原理。

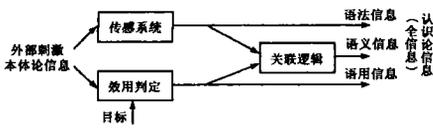


图 2 由本体论信息到全信息转换的原理

其中语法信息获取的原理(传感)一目了然,无需解释;语用信息 I_p 和语义信息 I_s 的获取则分别需要进行式 (3) 的效用判定和 (4) 的逻辑关联操作:

$$I_p = W \cos(S, G) \quad (3)$$

$$I_s = a(I_s, C I_s) \quad (4)$$

式 (3) 是计算外来刺激(信息)矢量 S 与系统的目标矢量 G 之间夹角的余弦值: 大于零为正效用, 小于零为负效用。式 (4) 是关联逻辑运算, 它表明: I 具有语法信息所描述的形式且具有语用信息所描述的效用 0 就是与之相应的 I 语义信息 0 。于是, 图 2 所示的系统可以完成从本体论信息到认识论信息的转换, 即相关信息的获取。

4.1.2 转换 2 由认识论信息到知识(认知)

智能生成共性机制的第二个转换是由信息提炼知识。专家系统的研究曾经关注过知识的问题, 但是没有涉及如何从信息提炼知识的问题, 大多数专家系统的知识都是由系统设计人员手工完成的。20 世纪 90 年代以来兴起的 /Data Mining and Knowledge Discovery 0 关注了如何从数据中提炼知识的问题, 但尚未形成普遍性和系统性理论。

2000 年作者在 /知识理论框架 0^[20] 中给出了知识的概念、分类、描述和知识量的测度方法, 探讨了 /把信息提炼为知识 0 和 /把知识激活为智能策略 0 的方法。

作者认为, 某个事物的信息表现的是 /该事物运动的状态及其变化的具体方式 Q 而事物的知识表达的则是 /该类事物运动的状态及其变化的抽象规律 0 。由 /具体的变化方式 0 到 /抽象的变化规律 0 的过程, 正是从信息资源中提炼知识的过程。因此, 由信息到知识的转换原理, 本质上是一种归纳和抽象的处理过程。

同 /信息 0 有语法信息、语义信息、语用信息分量的情形相对应, /知识 0 也有 /形式性知识 0 、/内容性知识 0 、/价值性知识 0 三个分量。其中, 事物的形式性知识反映的是事物结构形态方面的知识, 可以用结构/关联性 L 0 作为描述参量; 内容性知识反映的是事物逻辑涵义方面的知识, 可以用逻辑/合理性 R 0 作为描述参量; 价值性知识描述的是事物对于主体所呈现的价值信息, 可以直接用 /价值 V 0 作为描述参量。

因此, 对于一类事物 X , 如果它具有 N 种可能的运动状态, 在知识理论框架下, 与它相应的知识可以描述如下:

$$\begin{bmatrix} X_1, \dots, X_n, \dots, X_N \\ L_1, \dots, L_n, \dots, L_N \\ R_1, \dots, R_n, \dots, R_N \\ V_1, \dots, V_n, \dots, V_N \end{bmatrix} \quad (5)$$

其中, $0f(\cdot), [1, P, n, 2(\cdot)]_n > = < 1, (\cdot)$ 分别代表 L, r, v , 其

他符号意义同前。

由于篇幅原因, 这里不能引述知识理论的全部结果。有需要的读者可以参阅文献 [20]。

对照信息和知识的定义, 可以启示从信息提炼知识的基本原理。如前所说, 事物的信息是关于事物运动状态及其变化方式的表述; 事物的知识是认识主体关于事物运动状态及其变化规律的表述。因此从信息提炼知识的过程, 就是 /从具体现象到抽象规律 0 的归纳过程, 即从同类事物的大量具体 /状态变化方式 0 经过归纳处理, 抽象出反映其中 /状态变化规律 0 的过程。它的基本原理如图 3 所示。

当然, 依据不同的问题和约束条件, 具体的归纳和抽象算法会有所不同。但是, 归纳和抽象的原则是普遍的。事实上, 目前文献中关于数据挖掘中的各种算法都可以看作是归纳算法在各种不同情况下的具体实现。

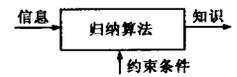


图 3 由信息提炼知识的归纳过程

一般来说, 事物的形式性知识可以从大量同类事物的语法信息归纳抽象出来, 事物的内容性知识可以从大量同类事物的语义信息归纳抽象得到, 事物的价值性知识则可以从大量同类事物的语用信息归纳抽象而成。但是, 同信息的情形 (4) 类似, 内容性知识的归纳需要与形式性知识和价值性知识共同作用, 如图 4 所示。



图 4 由信息到知识的转换

需要注意的是, /知识 0 并不是一个静止的概念, 它有自己的生长过程。知识的最初形态是经验, 这是一种具有粗糙真理性但是还没有经过严格的科学验证的情形, 可称为 /经验性知识 0 。科学验证为真的经验性知识, 称为 /规范性知识 0 , 是知识生长的第二阶段。规范性知识经过普及成为 /不证自明的公理 Q , 称为 /常识性知识 0 。这是知识生长的第三阶段。当然, 并非所有规范性知识最终都可以成为常识性知识; 另一方面, 那些老幼皆知和 /不言而喻的共识 0 以及本能性知识也应归入常识性知识的范畴。图 5 表示了知识的生长阶段^[17]。

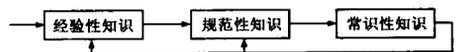


图 5 知识的生长阶段

知识理论的研究近几年受到越来越多的重视, 正在取得可喜的进展。限于篇幅, 关于这方面的讨论不在这里展开。有兴趣的读者可以参考相关的文献。

4.1.3 转换 3 从知识到智能策略(决策)

智能生成机制的第三个转换是由知识到策略的转换。由于策略比较集中地体现了求解问题的智能, 因此也常常

把它称为/智能策略0。当然,准确地说,完整的智能概念应当包含智能策略生成的过程以及智能策略应用的过程。所以,策略体现的其实只是狭义智能。

生成智能策略的重要条件是要具备相关问题及其环境的足够知识和信息;生成智能策略的另一个重要条件是要有明确的目标。前者为生成智能策略提供必要的基础,后者为生成智能策略提供引导的方向。基础和方向,两者缺一不可。可以认为,与其他问题不同,求解智能策略的一个重要重要特色就是/目标导引0。没有目标,就谈不上智能。因此,生成智能策略的过程实质上就是在给定/问题及其环境的知识和信息以及求解目标的信息0的约束条件下求解问题的过程,它的原理如图6所示。

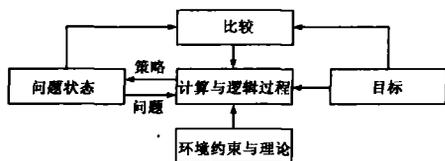


图6 由知识到智能策略的转换

图中示出,/智能策略生成0的机制是:(1)按照问题、环境、目标的约束,根据相关理论知识,通过计算和逻辑处理产生初始策略;(2)把初始策略作用于问题,使问题的原有状态改变为新状态;(3)将问题的新状态与目标状态进行比较:如果新状态与目标之间的差异比原有状态与目标之间的差异小,就按照原有的计算与逻辑处理继续前进,产生策略的后续部分,直到问题最新状态与目标之间的差异足够小,表示问题得到了满意的解决。这时得到的策略就是既能满足环境约束、又能解决问题、达到预期目标的/智能策略0;(4)否则就要回到步骤(1),在给定的知识和信息驱动下以及在目标的引导下改变原先的计算与逻辑处理,产生新的初始策略。

不过需要注意,在给定/问题2环境的知识和信息0的条件下,图中示出的这个求解过程可能有解也可能无解,取决于设定的预期目标是否合理。如果出现无解的情形,就只能退而求其次:或接受非最优解;或者需要修订目标重新求解。有时预期目标本身是合理的,然而给定的/问题及其环境的知识和信息0不够充分,也会导致不满意的求解结果。在这种条件下就需要设法得到更充分的知识和信息,否则就只能接受非优的求解结果。

一般,给定/问题2约束的知识和信息以及预设目标0之后,图7所示的求解智能策略方法原则上是可行的。不过,由于所利用的知识处于不同的生长阶段(见4.2),这个一般性的原理将会有不同的具体实现方式(见第5节的图7)。

4.1.4 转换4 由智能策略到智能行为(执行)

生成求解问题的智能策略之后,后续的过程就是要执行这个智能策略,即把智能策略转换为智能行为,使实际问题得到真正的解决。从功能的意义上说,控制系统就是

完成由智能策略到智能行为转换的技术系统。由于这部分内容比较熟悉,这里不再展开。

5 机制主义:结构主义、功能主义、行为主义的统一

机制主义认为,/信息2知识2智能转换0是智能生成的共性核心机制^[17]。尽管存在各种不同的智能生成过程,但是/信息2知识2智能转换0却是智能生成过程的共同主轴,差别主要表现为这些转换的具体实现方式的多样性。根据上节的讨论,可以得到图7的结果。

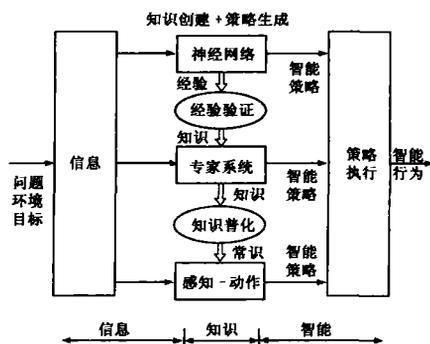


图7 信息2知识2智能转换:智能生成的共性核心机制

图7表明,结构主义方法(利用经验性知识的神经网络,其中也涉及到模糊逻辑方法和粗糙集合方法)、功能主义方法(利用规范性知识的专家系统方法)、行为主义方法(利用常识性知识的感知2动作系统方法)三者确实在机制主义(信息2知识2智能转换)框架下实现了完美的互补与统一。而且,结构主义方法获得的经验性知识经过验证就可以成为功能主义方法所需要的规范性知识,功能主义方法的规范性知识经过普及处理就可以成为行为主义方法所需要的常识性知识。

可见,在智能生成机制(机制主义)的统一框架体系下,结构主义(以人工神经网络为代表)、功能主义(以专家系统位代表)、行为主义(以感知动作系统为代表)三者之间并不存在/孰优孰劣0的分别;相反,它们之间构成了和谐分工互补的统一体。于是,/智能生成的共性机制2信息2知识2智能转换0就天然地成为了/人工智能统一理论0的基础。

作者相信,/人工智能统一理论0不仅有助于结束原来三种理论之间旷日持久的/孰优孰劣0之争;而且可以使人工智能的整体理论由此得到有效的深化和强化,为未来的发展创造新的基础,提供新的机会。

参考文献:

[1] W C McCulloch, W Pitts A logic calculus of the ideas immanent in nervous activity[J]. Bulletin of Mathematical Biophysics 1943, 5: 115- 133.
 [2] Rosenblatt F. The Perceptron: A probabilistic model for information storage and organization in the brain[J].

- Psych Rev 1958 65: 386- 408
- [3] Widrow B M E Hoff Adaptive switching circuits[A]. 1960 IRE Western Electric Show and Convention Record [C]. 1960 4(23). 96- 104
- [4] Amari S I Neural theory of association and concept formation[J]. BiolCybern, 1977, 26: 175- 185.
- [5] Hopfield J J Neural networks and physical systems with emergent collective computational abilities[J]. Proc Natl Acad Sci 1982, 79: 2554- 25558
- [6] Grossberg S Studies of Mind and Brain Neural Principles of Learning Perception Development Cognition and Motor Control[M]. Boston: Reidel Press 1982
- [7] Kihonen T. Self-Organization and Associative Memory [M]. Berlin: Springer Verlag 1984
- [8] Rumelhart D E Brain style computation: learning and generalization[A]. in Introduction to Neural and Electronic Networks[C]. New York: Academic Press 1990
- [9] Newell A, Simon H A GPS: A program that simulates human thought[A]. in Computers and Thought[C]. (see [10]) McGraw-Hill New York, 1963. 279- 293
- [10] Feigenbaum E A, J Feldman (Eds). Computers and Thought[M]. McGraw-Hill 1963.
- [11] Simon H A. The Sciences of Artificial[M]. Cambridge MA: The MIT Press 1969
- [12] Newell A, H A Simon Human Problem Solving[M]. Englewood Cliffs NJ: Prentice-Hall 1972
- [13] McCarthy J H History of LISP[A]. SIGPLAN Notices[C]. 1978, 13: 217- 223
- [14] Minsky M L. The Society of Mind[M]. Simon and Schuster New York, 1986
- [15] Brooks R A. Intelligence without representation[J]. Artificial Intelligence, 1991, 47: 139- 159
- [16] Brooks R A. Intelligence without reasoning[A]. in Proceedings of IJCAI 1991[C]. Sydney, Australia 1991: 569- 595
- [17] 钟义信. 知行学引论))) 信息和知识智能转换理论 [J]. 中国工程科学, 2004 6: 1- 8
Y X Zhong. An introduction to cognetics: theory of transformations from information to knowledge and further to intelligence[J]. Chinese Engineering Science, 2004 6: 1- 8 (in Chinese)
- [18] C E Shannon Mathematical theory of communication[J]. BSI J 1948, 27: 379- 423, 623- 656
- [19] 钟义信. 信息科学原理 [M]. 北京: 北京邮电大学出版社, 2002
Y X Zhong. Principles of Information Science[M]. Beijing: BUPT Press 2002 (in Chinese)
- [20] Y X Zhong. Framework of knowledge theory [A]. Proceedings of SCI[C]. Orlando, Florida, USA, 2000: 232- 265, 531- 542

作者简介:

钟义信 男, 1940年 2月出生于江西龙南, 北京邮电大学教授, 近期主要研究方向: 信息科学, 人工智能. E-mail: yxzhong@ieee.org