

国家自然科学基金资助项目

认知科学与 广义进化论

赵南元 著



清华大学出版社

375617

认知科学与广义进化论

赵南元 著

国家自然科学基金资助项目



清华大学出版社

(京)新登字 158 号

图书在版编目 (CIP) 数据

认知科学与广义进化论/赵南元著. -北京:清华大学出版社,
1994

ISBN 7-302-01446-9

I. 认… II. 赵… III. ①认识论-广义进化论②广义进化论
-认识论 IV①B017②Q111

中国版本图书馆 CIP 数据核字 (94) 第 01373 号

出版者:清华大学出版社 (北京清华大学校内, 邮编 100084)

印刷者:北京密云胶印厂

发行者:新华书店总店北京科技发行所

开 本: 850×1168 1/32 印张: 11.75 字数: 303 千字

版 次: 1994 年 3 月第 1 版 1994 年 3 月第 1 次印刷

本社分类号: N·1

印 数: 0001—4000

定 价: 16.80 元

内 容 简 介

《认知科学与广义进化论》一书是著者所承担的国家自然科学基金项目的研究成果。在本书中著者创立了横跨哲学认识论、比较文化学、进化生物学、计算机科学、系统科学等多学科的高度统一的认知科学基础理论——广义进化论。并探讨了该理论在建立伦理学、美学等社会科学基础方面的作用。本书论点丰富新颖，论证严密科学，表达深入浅出，不仅适合于有关学科的教学、科研人员参考，而且适合于一般读者阅读。

序 言

本书的目标是建立认知科学的核心理论。通俗地说，就是要回答这样的问题：人的大脑（或心灵）是按照什么原理工作的？人工智能的研究曾经试图解决这个问题，但是传统人工智能在这一方面显然没有能够取得令人满意的结果。本书的第一章就是分析这个问题的困难之处，为读者提供必要的背景知识。人工智能的失误并不是技术上的失误，而是哲学层次上的失误。为了解决这个世纪性的难题，我们不得不首先从哲学层次上建立与我们的目标相适应的理论，即工程认识论。这种认识论与通常的基于本体论的认识论的区别，根植于东西方文化的深层精神结构的不同，因此我们在第二章从比较文化学的角度入手，从更深层去挖掘认识论差异的根源。认识论的差异导致了对科学本身看法的变化，因此在第二章里我们也讨论了一些与科学哲学有关的问题。

第三章是本书的核心理论——广义进化论。把广义进化论作为认知科学的基础，体现了一种被很多认知科学的研究者所忽视的事实，即人是一种动物，人是从其他动物进化而来的，人类的认知和思维能力也是在进化之中产生的。另一方面，人与机器的最大区别是人具有创造性，而为了理解创造过程，最好的方法莫过于借鉴自然界中最为壮丽的创造过程——进化。广义进化论是对达尔文进化论的发展。当我们用新的眼光观察生物界时，会产生新的观点。在第三章中我们用广义进化论的观点讨论了进化生物学的一些重要问题，展示了广义进化论的理论威力。在第四章中，我们运用广义进化论的观点，对认知和思维中的重大问题进行了分析，建立了感情和意识的具有工程可实现性的理论模型，阐

• ■ •

述了感情和意识在认知中的作用和机制。

创造源于进化，进化的基本机制是变异与选择，选择的根据是价值评价。因此本书在哲学层次上建立了关于价值的基础理论，并将这一精神贯彻于全书内容之中，成为全书的一条主线。为自然科学与社会科学寻找共同的理论基础，是科学对于深刻性的必然追求。由于本书建立了关于价值的深层理论、关于进化的广义理论、以及关于人类决策的认知理论，也就提供了自然科学与社会科学的共同基础。以往社会科学建立基础理论的困难主要是善恶评价的主观性、社会不断发展变化的自我表述性、以及作为社会的基本单元的人的复杂性。本书的理论对于克服这些困难是有帮助的。在第五章里我们用伦理学、美学等实例，探讨了本书的理论在社会科学和复杂巨系统理论等领域中的作用。

由于本书所涉及的学科范围很广，不可能要求读者（包括著者）在所有这些领域都是专家，而且也考虑到非专业人员阅读的可能性，所以本书的写作方针是“深入浅出、点到为止”。在文字表达上力求通俗，尽量避免数学形式的表达。对于“好读书，不求甚解”，仅以扩大知识面为目的的读者来说，本书的可读性也是很强的，足以提供茶余饭后之谈资。对于想要深入理解本书内容的人来说，需要多下些工夫才行。本书所研究的问题本身是很困难的，新概念比较多，大多数概念要通过与其他概念的相互关系才能深入理解，因此建议读者在阅读本书时可以采用多次反复阅读的方法，不清楚之处先放过去，等下次再读时，可能就容易理解了。可以说，只阅读本书的个别章节是很难理解的，任何一个章节的深入理解都需要整体的支持，这就是反复阅读的理由。此外，由于本书在有限的篇幅内容纳了庞大的理论体系，使得叙述趋于精炼，有些地方会有说明不充分之感，甚至有些跳跃。其中有些是在其他有关学科中已经研究过或正在研究的内容，读者可以参考有关文献以加深认识，更多的则是尚未深入研究的部分，可

以作为未来研究的课题。本书的贡献不仅是建立了一个理论框架，更重要的是开辟了广阔的研究领域。本书的一节，乃至一个自然段所讨论的问题，有不少可足以构成一个博士论文的题目，有待读者进行发掘。著者在清华大学开设“认知科学导论”的研究生课程，主要讲授本书的内容。本书可以作为认知科学的教科书使用。

感谢国家自然科学基金委员会对本研究项目提供的资助，为本项目创造了宽松的研究环境。

感谢日本的株式会社 *Ampere* 及社长草椰高志先生为著者在日本为期两年的研究活动提供了充足的费用和良好的工作及生活条件，在这段时间的研究中本项目取得了很大的进展。

感谢已故的常迥教授。如果没有常迥教授对本研究项目的极力支持和鼓励，著者不可能有勇气进行如此艰巨的挑战。常迥教授作为项目总负责人未能在生前见到本书的完成，是著者最大的遗憾。

感谢著者在攻读博士时的导师饭岛泰藏教授，正是他的影响使著者从一个工程技术人员转变为对基础理论抱有强烈兴趣的科学工作者。

感谢王文渊教授在本项目研究过程中与著者的充分合作。为使著者可以集中精力进行理论研究和撰写本书，他领导了本项目中所有实验研究，承担了项目组的所有日常具体工作，还接替了著者担负的行政工作，并对著者的工作不断提出积极的建议。本书能够按期完成，与王文渊教授的贡献是分不开的。

感谢李衍达教授对本书进行了审阅和出版推荐，并对本书内容提出深刻的意见，使本书进一步完善。同时感谢本书责任编辑在本书的编辑出版工作中所付出的辛勤劳动。

感谢边肇祺教授、阎平凡教授长期以来对本项研究的全力支持和热情关怀，以及在学术方面起到的推动作用。

参与本项目研究的李小平博士、刘力夫博士、胡胜发博士、以及博士生范列湘、硕士及硕士生丁志敏、蔡奕、于婷、李庆华、黄兵、王嘉欣等，作了大量的研究工作或提出了自己的看法，为本项目作出了各自的贡献，特此致谢。

在研究过程中，林行刚教授、杨福生教授、陈霖教授、马颂德教授、迟惠生教授、马希文教授、小川英光教授、佐藤诚教授、中野馨教授、黄煦先生曾与著者就本项目有关的问题进行过有益的讨论，特此致谢。

著者所在单位的各级领导、同事及著者的家庭成员为此项研究创造了良好的环境，特此致谢。

在本书撰写过程中，钱学森教授来信对本书表示关心、肯定和支持，对本书的完成起到了巨大的促进作用，在此深表感谢。

北京清华大学自动化系 赵南元

1994年1月

目 录

序 言	(Ⅲ)
第一章 绪论	(1)
1.1 三个许诺和一则寓言	(1)
1.2 人工智能中的问题及其困难之处	(2)
1.3 认知科学的方向.....	(17)
第二章 工程认识论	(41)
2.1 东西方文化比较和深层精神结构.....	(46)
2.2 科学与形而上学.....	(61)
2.3 从真理到善理.....	(69)
2.4 认识论的视镜模型.....	(96)
第三章 广义进化论	(113)
3.1 自我表述系统	(113)
3.2 进化的地形图模型	(117)
3.3 进化论的博弈论方法	(122)
3.4 广义进化论的软硬结构模型	(135)
3.5 生物的进化与进步	(151)
3.6 有性生殖的意义	(165)
3.7 广义进化范式	(178)
第四章 认知与思维	(183)

4.1	认知的阶层性	(183)
4.2	神经网络	(190)
4.3	神经网络群体	(203)
4.4	从动物到人的遗传与学习	(209)
4.5	智能模型与形象思维	(231)
4.6	感情与评价	(250)
4.7	意识与自我	(270)
4.8	创造与灵感	(291)
4.9	模式识别	(301)
4.10	人工认知主体	(304)
第五章 广义进化系统		(309)
5.1	复杂巨系统	(309)
5.2	全局观点与分离主义	(313)
5.3	伦理学的非形而上学基础	(326)
5.4	美学的情绪基础	(352)
5.5	社会的通用评价尺度——论金钱	(356)
参考文献		(364)

第一章 绪 论

1.1 三个许诺和一则寓言

大约 40 年前，正值计算机科学的黎明时期，计算机专家许下了三个宏伟的诺言：

10 年之后

- ① 计算机可以发现有价值的新定理。
- ② 计算机可以夺得国际象棋的世界冠军。
- ③ 大部分心理学理论将采用计算机程序的形式。

40 年后的今天，这些诺言还没有一个是完全实现了的。这 40 年间，计算机硬件得到了飞速的发展，计算速度、存储容量、性能价格比、性能重量（体积）比等各项参数都按指数规律增长。如果这个增长势头能保持下去，50 年后，计算机的单位重量的计算能力和存储容量都可以超过人的大脑。但是 50 年后我们能不能用这种高能力计算机作为大脑制造出象科幻影片中出现的那种具有人的能力的机器人呢？对于这个问题，恐怕现在的人工智能研究还难以给出肯定的回答，我们所掌握的关于大脑认知过程的知识还差得很远，即使硬件条件具备，软件和制作软件的理论可能还达不到实现这一目标的要求。

对于这种现状，一位学者作了极其生动的比喻，讲了这样一个寓言：“警察问一个在路灯下找东西的人：‘你在找什么呢？’‘我在找钥匙’‘你的钥匙是丢在这儿了吗？’‘不是’‘那你为什么要在这儿找呢？’‘因为这儿比别处亮’”这个寓言本来是针对计算机视觉的研究现状而言的，但是看起来也符合整个人工智能领域

的现状。在这个领域里，研究者往往忽视计算机与人脑的不同之处，热衷于搞一些利用现成的数学工具及其他理论工具所容易解决的问题，结果是研究出的方法与实际的视觉及智能的机制相去甚远。

为了改变这种现状，需要解决的问题是：

- ① 钥匙到底丢在哪儿了？
- ② 那儿为什么暗？
- ③ 能不能让暗处变亮？

这三个问题换一种说法就是：

- ① 人工智能的根本性问题在何处？
- ② 解决这些问题存在什么困难？
- ③ 这些困难能不能克服？

以下我们先就人工智能的一些领域来分析一下上述①②两个问题，而第③个问题则是本书整个篇幅的中心内容。

1.2 人工智能中的问题及其困难之处

1. 指数爆炸

中国古代关于指数爆炸的思考见于“歧路亡羊”的故事：某人的羊丢了，他请村里的人一起去找，不久他愁眉苦脸的回来了，别人问他：“不就是一只羊么，何必如此沮丧。”他说：“不是羊的问题，我想的是，路上有岔路，岔路前头还有岔路，这样算起来村里有多少人也是不会够的”。当然在日常生活中羊是可能找到的，这是一个思考实验，显示了指数爆炸的困难。

外国也有这样的故事。一个国王非常喜欢国际象棋，就招见发明国际象棋的人问他想要什么奖赏，他在棋盘的第一个格里放了1颗麦粒，在第二个格里放2颗，第三个格放4颗，说道：按

这样放下去，直到放满棋盘的 64 个格子。大臣们计算之后发现，要拿出这么多麦子，这个小小的王国无论如何是办不到的。 2^{64} 颗麦子够现在全世界的人吃几十年！

早期人工智能研究的一个中心课题是象下棋这样的游戏。这是因为在当时计算机的发明者著名数学家冯·诺伊曼 (Von Neumann) 已经建立了关于游戏的理论基础“博弈论”，而下棋又属于博弈论中最简单的一种情况“二人零和博弈”。二人是指只有两个人参加游戏，零和是说一方胜了则另一方为负，没有两人都胜的情况，一方的所得就是对方的所失，两人加起来得失之和为零。玩零和博弈的基本原则是“极大极小准则”，即要选择一种策略力求使自己的最小收益最大化。用一个简单的例子来说，假定我有 AB 两种策略，用 A 策略对方如果不注意会被我将死，但如果注意到会吃我一个车，而如果用 B 策略，则无论对方怎么下我都能吃他一个卒。按极大极小准则我应该选择 B 策略，而不应用 A 策略去下险棋，不能假定对方是个笨蛋。按照这个准则，再规定一组评价函数，例如吃一个车得 10 分，丢一个马得 -6 分等等，把这些原则放到计算机里去，计算机就会对所有可能的走法算出分数，按极大极小原则选出最稳妥的一步棋。到此为止没什么大问题，问题出在后头。显然，只看一步棋是不行的。对手可以先给你点甜头让你上当，然后把你杀个落花流水。为了取胜必须多看几步。假如对于每个棋局可走的方法有 100 种，对于每一种走法对方也有 100 种走法，看双方一步棋要算一万次，这对计算机来说不算多，可是要看双方两步就要算一亿次，用每秒一亿次的计算机要算一秒钟，如果看三步就得三个小时，看四步得三年，这棋就下不成了。

那么能不能用制造更快的计算机的办法来解决这个问题呢？回答是否定的。因为指数规律的增长实在太快了，而计算机受物理定律的制约，不可能作到无限大和无限快，速度受光速的限制，

元件的缩小受测不准原理的限制,元件的个数受地球资源的限制。根据勃瑞姆曼极限(陈禹:《关于系统的对话》)的计算,一个由最小的元件构成的相当于地球质量的最快的计算机,从地球诞生时一直计算到现在,可以处理的信息量小于 10^{93} 位(bit,信息量的最小单位)。 10^{93} 看起来相当大,但是指数增长却可以轻而易举的突破它。例如一个围棋盘有 $19 \times 19 = 361$ 个点,每个点可以有黑子、白子、没子的三个状态,那么所有可能的棋局就是 $3^{361} > 3^{194} \approx 10^{93}$ 。看来与指数爆炸发生正面冲突就象鸡蛋碰石头。

那么能不能想出一个巧妙的算法来减少计算次数呢?这个问题涉及计算复杂性这一专门研究领域。一般认为,一类问题的计算复杂性是问题本身的性质,与算法无关。而且人的大脑也同样不能解决指数爆炸的问题。由此可以得到一个重要的结论:如果我们对大脑所作的模型遇到指数爆炸,必然说明这个模型与大脑中的实际过程相差甚远,需要重新考虑建立新的模型。换句话说,指数爆炸是提醒我们已经走错了方向的极好的警告灯。当然,指数爆炸的困难不是绝对的,如果问题的规模比较小,就能够在可以容忍的时间里得到解决。例如有一种圈叉游戏,只有 $3 \times 3 = 9$ 个格子,一方在格子里画圈,一方画叉,看谁先把自己的记号连成三个一排,这种游戏只有 $3^9 = 19683$ 种棋局,用计算机下这种棋是很容易的。只是这种游戏没什么意思。

游戏的复杂程度差别很大,最简单的游戏是有必胜法的游戏,只要游戏双方都掌握了这种法则,并完全照办,那么从游戏一开始就注定了是先走的胜还是后走的胜。如果这种必胜法不太简明扼要,人记起来比较困难,那么机器就容易胜过人了。例如三堆火柴游戏就属于有必胜法的游戏。游戏的规则是:放三堆火柴,双方轮流从任意一堆中拿走任意根火柴(至少拿一根),拿最后一根的人算输。这种游戏的解空间不大,只有三维。如果每堆火柴不超过一百根,全部解只对应于三维空间的一百万个点,可以事先

从小到大算出这些点哪些必胜，哪些必败。如果一开始的火柴数在败点上，先拿的人必输，只能等待对手出差错。如果一开始是胜点，则只要把火柴数拿到败点上交给对方，先拿的一方稳操胜券。一开局就已知胜负的游戏，当然没多大意思。发现了必胜法，等于对那种游戏判了死刑。

象棋和跳棋之类则找不到必胜法。好在其近期评价（吃子，前跳）与最终评价（将军，跳完）之间比较一致，一般来说吃子可以增加将军的机会，因此用能力较大的计算机还可以得到不错的结果，甚至能打败世界冠军，但是还显得有些死板，世界冠军多下几次就会摸出计算机的脾气，把计算机打败。而围棋则要复杂得多，落子的近期效果与远期效果未必一致，评价准则也很不明确，所以计算机下围棋的程序至今也还没有达到一段的水平。

早期的人工智能从玩游戏和证明定理出发，希望找出能解决所有问题的通用方法来，即开发“通用解题者”的程序。这一方面遇到了指数爆炸的困难，另一方面又让人感到不符合人思考的实际情况。人在下棋时显然不会考虑几亿个可能的棋局，只能借助知识在不太多的有希望的走法中选择。出于这样的认识，60年代后期人工智能开始转向对“知识”的研究，希望找到一种通用的在计算机中表现知识和利用知识去解决问题的方法。在这个时期出现了很多知识表现手法，例如声明型知识（什么是什么）、过程型知识（怎样作什么）、知识框架、剧本型知识等等，并制成了专家系统，把专家的知识放在计算机的知识库里，用这些知识解决一些专门领域的问题。这个研究方向也遇到一些共同的困难，第一是专家的丰富知识并不全是说得清楚的，有一些知识近乎直觉，说不清的知识不能形式化，也就无法放到计算机里去。与此有关的另一个困难是，这些知识表现方法都不是面向学习的，即机器不能自己学习新的知识，而真正要作出能应付实用的专家系统，需要的知识是非常大量的。这些知识由人一条条写好送到机器里去，

其工作量有时会大得惊人，特别是考虑到人的日常生活中的知识（常识），其规模之大远远超乎一般人的想象。特别是除了严密的科学领域之外，一般的知识往往包含例外情况，而不考虑例外情况的专家系统一遇到例外就会出错。但考虑了例外则又会使知识极度膨胀，明斯基（M. Minsky）曾举了一个生动的例子来描述这种情况：“鸟能飞——如果不是企鹅或鸵鸟，如果还没死，如果翅膀没断，如果没有关在鸟笼里，如果脚没铸在水泥里，如果没有遇到过足以使其不能飞的心理上的恐惧经验。”^[1]显然这些还远没包括所有的例外因素，何况例外之中还有例外，岔路上还有岔路。

上述的研究方向统称符号主义的方向。符号主义方向在学习方面存在着根本性的困难，这是因为人或机器能够直接从客观世界取得的信息都是模式信息，例如文字、图象、语音等等，而模式信息的处理又在符号主义的视野之外，要想直接从客观世界中获得知识，只处理符号是不够的。当符号主义方向的局限性日趋明显之后，和符号主义的历史一样久远而一度受到冷落的连接主义方向重新获得了重视。

早在40年代冯·诺依曼提出了计算机的基本结构之后不久，就研究过逐次处理的计算机与人脑的区别，提出了神经元网络型计算机的想法。50年代末罗森布拉特（F. Rosenblatt）试制了感知机（*Perceptron*）相当于双层的神经网络，具备学习和处理模式的能力。60年代末，明斯基等人（Minsky, Papert）出版了一部名为感知机的专著，指出二层神经网络只能解决线性可分离的问题，解决不了异或（*XOR*）问题，而多层神经网络虽然能解决任意复杂的问题，却缺乏有效的学习算法。这个结论使得神经网络的研究一度冷却下来。到80年代，出现了一些关于神经网络的新算法，特别是Rumelhart和McClelland提出的误差反向传播算法（*Back Propagation*，简称BP算法），证明了多层网存在有效的学习算法，并

实际解决了包括异或问题在内的一批问题，使得关于神经网络的研究掀起了一个高潮。通过近年来的深入研究，神经网的局限性也越来越明朗化了，神经网络的学习时间随问题的复杂程度呈指数增长。这里同样存在着指数爆炸。例如如果让神经网络去通过示例学习区分质数与非质数的一般性方法是做不到的。值得注意的是，这种难度过高的问题人也解决不了，如果对一个会数数但还不懂乘除法的儿童提示大量的质数与非质数的实例，让儿童学到能判断一个未见过的数是不是质数，同样是作不到的。这使我们想起冯·诺伊曼在研究计算机与人脑的不同点时得到的重要结论：计算机可以处理逻辑深度很大的问题，而人则只能处理逻辑深度很小的问题。从这一角度上看，神经网的特性与人脑是相近的。但是如果把人脑看成是一个均匀的大神经网络，不仅不符合脑神经科学的现实，而且也会遇到数学上指数爆炸的障碍。大脑是由大量的小神经网络模块井然有序组织起来的大系统。要想知道低能力的神经网络怎样组成高能力的大脑，还必须知道更高层次的组织原则。这已经在神经网络的研究领域之外了。

符号主义与连接主义扎根于不同的哲学土壤。从认识论角度来看，符号主义属于先验论的哲学体系，连接主义则属于经验论。现在看来两者各有其片面性，但是两者简单的合取或折衷未必能产生更好的认识论，这里需要新的哲学体系。

2. 模式识别与丑小鸭定理

模式识别是人工智能的一个重要领域，其目标是把模式信息转换成便于计算机处理的符号信息。模式识别的处理对象范围很广，主要可以分为二维和一维模式两大类（三维以上的模式信息量太大，难以直接处理）。二维模式识别主要包括图象识别，文字识别，物体识别，纹理识别，景物识别，三维运动目标识别等等；一维则包括语音识别及各种波形识别。模式识别的过程大致可以