

# Herbert Simon 在最后几年里的两个研究方向

秦裕林<sup>1</sup>, 朱新民<sup>2</sup>, 朱 丹<sup>3</sup>

(浙江大学 理学院, 浙江 杭州 310027; 2. 中国科学院 心理研究所, 北京 100101; 3. 爱荷华州立大学 商学院)

**摘 要:** Newell, Shaw 和 Simon 的“逻辑理论家”(LT) 是 1956 年 6 月在 Dartmouth 研讨会上唯一的编出了程序的人工智能实例, 这个研讨会宣告了人工智能(AI)的诞生. 今天, 在纪念人工智能诞生 50 周年之际, 本文拟介绍 Herbert A. Simon 在他人生最后几年的研究领域中的 2 个方向——例中学和多种表达. 本文的笔者们曾作为 Simon 的学生、同事和朋友参加了这 2 个研究方向的工作, 谨以此文表达我们对 Simon (1916-2001) 的哀思及追念.

**关键词:** 例中学; 多种表达式; 人工智能

**中图分类号:** TP18 **文献标识码:** A **文章编号:** 1673-4785(2006)02-0011-05

## Herbert Simon's two research directions in his lost years

QIN Yu-lin, ZHU Xin-min, ZHU Dan

(1. College of Science, Zhejiang University, Hangzhou 310027, China; 2. Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China; 3. College of Business, Iowa State University, IA 50011, USA)

**Abstract:** Newell, Shaw and Simon's "logic theorist" (LT) was the only tangible (programmed) example of artificial intelligence (AI) at the Dartmouth workshop in June 1956, which declared the birth of artificial intelligence and was the central event celebrated at this meeting. In this article, we discuss two of Herbert Simon's research directions during his last years—learning from examples and multiple representations. The authors were involved in these researches as his students, colleagues and friends. This paper was dedicated as a memorial to Herbert Simon (1916-2001).

**Key words:** learning from examples; multiple representation; artificial intelligent

Herbert A. Simon (1916 - 2001) 生前系美国卡内基梅隆大学的计算机系和心理学系教授, 人工智能理论奠基人之一, 中国科学院首批外籍院士. 曾获得 1978 年诺贝尔经济奖, 1975 年图灵奖 (与 Newell 共享), 1995 年国际人工智能协会杰出研究奖, 以及 1986 年美国国家科学金奖, 1993 年美国心理学会终身贡献奖等重大奖项. 他在美国卡内基梅隆大学的计算机系和心理学系的网页至今还保存着. 在计算机系的网页上 (<http://www.psy.cmu.edu/psy/faculty/hsimon/comp-sci.html>) 这样介绍: “Herbert Simon 在计算机方面的主要兴趣是人工智能、人机界面、人和机器的组织原理是信息加工系统, 用计算机 (通过建立模型) 来研究有关智能本质以及认识论的哲学问题, 以及计算机技术的社会

意义 (social implications).

40 余年来, Simon 从事于建立模拟人的不断增加复杂性的处理, 非良构的 (ill-structured) 认知任务的过程的计算机模拟模型, 其中包括科学发现过程和思维中视觉图像的使用. 更多细节见心理学系网页, 在心理学系网页上 (<http://www.psy.cmu.edu/psy/faculty/hsimon/hsimon.html>) Simon 说道: “人类应用符号加工来解决问题, 推理, 说话, 写作, 学习与发明创造. 过去 45 年来, 认知心理学已经建立和检验了这些加工过程的经验模型. 开始是较简单的任务, 然后到较复杂的任务. 这些模型的形式是模拟人的行为的计算机程序.”

研究已经从这些解释人类思维的领域扩展到新的现象和领域, 主要是例中学, CaMeRa (一种用视觉图像推理的模型), 发现好的问题表达, EPAM (一

收稿日期: 2006-08-28.

个模拟知觉和记忆的统一理论),科学发现的心理学(BACON 及其他程序)。

文中仅介绍 Simon 他在人生的最后几年,在人工智能和认知心理学方面的众多研究方向中的2个研究方向——例中学及多种表达(CaMeRa)。本文的笔者参加了这2个研究方向的工作。

## 1 例中学(learning from examples)

Newell, Shaw 和 Simon 的“逻辑理论家(LT)”始于1955年,完成于1956年。这是在宣布了人工智能诞生的1956年6月Dartmouth研讨会上唯一的编了程序的AI实例<sup>[1]</sup>,LT对AI的主要贡献,除了LT本身以外,是启发式搜索<sup>[2]</sup>和表处理语言(命名为LPL(信息加工语言))<sup>[1]</sup>。

在1957年开发的通用问题解决者(GPS),他们给AI增加了手段——目的分析方法。这个方法是从分析测试者在实验室执行问题求解任务时的口述报告中发现的<sup>[3]</sup>。

在Newell 和 Simon 1972年出版的《人类问题求解》一书中<sup>[4]</sup>,他们申明人类问题求解的过程是调用产生式(条件—动作对)在问题空间中搜索从起点到目标的通道的过程。同样,这也是从分析测试者的口述报告中得来的。从这以后,产生式系统就成了Simon, Newell (如Soar<sup>[5]</sup>)的问题求解系统,以及许多其他问题求解系统(如John A Anderson 的ACT-R)的核心组成部分。

早在1955年,当还在开发LT时,Simon就考虑了能够学习和编程序的程序。在产生式系统的框架下,学习,至少是技能的学习,意味着新的产生式的建立。1978年,自适应产生式系统(APS),一种能够修改自身的产生式的产生式系统,被应用于计算机学习解代数题<sup>[6]</sup>。

朱和Simon在1987年的试验,“提供了坚实的证据,表明了通过呈现给学生仔细选择的排好解题步骤的例题和习题从而教会学生不同的数学技能的可能性。这中间不需要老师的讲课或其他的直接指导方法”<sup>[7]</sup>。

他们所用的例子之一是因式分解,对刚刚学了多项式乘法的初中生,指出因式分解是2个一次多项式的乘积的逆运算,同时给出一个例子, $x^2 + 5x + 6 = (x + 2)(x + 3)$ ,学生就能够学到新的产生式:

如果:目标是因式分解  $x^2 + ax + b$

那末:找2个整数  $c$  和  $d$ , 使满足  $c + d = a$ ,

$c \times d = b$ :

写下  $x^2 + ax + b = (x + c)(x + d)$ 。

这个学习过程,后来成功地用基于解释的泛化学习(Explanation-based generalization, EBG)方法<sup>[8-9]</sup>在THEO系统<sup>[10]</sup>上进行了模拟。下面是这个模拟所用的知识表达(由Prolog语言翻译成中文):

1) 目标概念(因式分解定义,因式分解是多项式的乘积的逆运算):

如果:  $(x + c)$  和  $(x + d)$  的乘积是  $x^2 + a + b$ ,

那末:  $x^2 + ax + b$  的因式是  $(x + c)$  和  $(x + d)$ 。

2) 领域知识(domain knowledge)

如果:已知整数  $c$  和  $d$ , 满足  $c + d = a, c \times d = b$ 。

那末:  $(x + c)$  和  $(x + d)$  的乘积是  $x^2 + ax + b$ 。

3) 实例

$x^2 + 5x + 6$  的因式是  $(x + 2)$  和  $(x + 3)$ 。

THEO可以由此形成因式分解的产生式规则,它类似于人类被试学得的因式分解产生式规则<sup>[11]</sup>。

Simon一直有一个想法,认为产生式的左边(条件部分)是学习形成新产生式的更重要的部分,而这一点恰恰为很多老师和教材所忽略(他们通常过分强调了学生掌握产生式的右边,即动作部分)。

1996年,朱新民、刘亦菲、Herbert Simon 和朱丹指出,通过呈现给学生仔细选择的排好解题步骤的例题和习题,能使学习注重学习产生式的条件部分,抓住相关线索,忽略无关信息<sup>[12]</sup>。

例如,在下面这两个题中,除了第一个条件外,其余都非常相似。

问题1 如果  $A$  和  $B$  均浮在水上,  $A$  比  $B$  重,但  $B$  比  $A$  体积大,假如  $A$  所受的浮力为  $F_b(A)$ ,  $B$  所受的浮力为  $F_b(B)$ , 以下断言中哪一个正确?

A.  $F_b(A) > F_b(B)$ ;

B.  $F_b(A) = F_b(B)$ ;

C.  $F_b(A) < F_b(B)$ ;

D. 所给条件不足以做出判断。

问题2 如果  $A$  和  $B$  全部浸没于水中,  $A$  比  $B$  重,但  $B$  比  $A$  体积大,假如  $A$  所受的浮力为  $F_b(A)$ ,  $B$  所受的浮力为  $F_b(B)$ , 以下断言中哪一个正确?

E.  $F_b(A) > F_b(B)$ ;

F.  $F_b(A) = F_b(B)$ ;

G.  $F_b(A) < F_b(B)$ ;

H. 所给条件不足以做出判断.

在这个实验中,学生通过学习能够和专家一样,认出在这2题中,第一个条件是关键因素.在问题1中,关键因素是“如果A和B均浮在水上”,因此,A重于B是有关信息,B大于A是无关信息.因而正确答案是A.在问题2中,关键因素变成了“A和B全部浸没于水中”,于是,A重于B成了无关信息,B大于A成为有关信息,因而正确答案是C.可以预测,学生有能力建立子目标来正确的回答下面这个更能“困惑”人的问题,“如果A和B都浮在水上,A比B重,但B比A体积大,哪一个物体排开的水更多?”尽管B比A体积大,学生仍然能正确地回答A排开的水更多.

更重要的是,Simon热切地想看到这些发现被用于现实世界中的学校教学过程中去,文献[13]综述了例中学,例中学用于教学的基本原理.令人鼓舞的结果表明,认知理论已经发展到能为实际教学提供有用的启发式信息(heuristics)的地步.它显示了从自适应产生式理论获得的经验,如何为实现有效的理解性学习提供了具体的实践指导.Simon在他的网页上总结道:“能够通过检查已解出的问题的例子来学习新技能的计算机程序已经建立,以这个程序为模型,经中国科学院的同事们的努力,中国的全部中学数学(代数,几何)教材已经设计出来,并且在二十几个中学获得了成功的应用.”

## 2 多种表达(multiple representation)

在1955年的一段时间,Simon想用计算机来模拟人求解几何题.但遇到了处理几何图形的困难.为了避免这个问题,他和Newell改选了“逻辑理论家”作为用选择性搜索的方法来解决非数值算计问题的第一个程序,从这以后,表达(representation)一直是他的研究领域中的核心问题之一.

1987年,Jill Larkin和Simon在他们的著名的“为何一幅图(有时候)能抵万个词”论文中<sup>[14]</sup>仔细地定义了表达的信息等价与计算等价,指出了一个表达的計算的有效性取决于它的数据结构,操作数据结果做出新推断的程序的的管理,以及它们是如何一起工作的.该文具体揭示了为什么图形表达有时会优于语句表达.

1995年在“图形推理”一书的前言中,Simon写道:“在一个全神贯注于信息高速公路的社会中,对图形推理的深刻理解将是保证这高速公路上车辆运行的基本条件”<sup>[15]</sup>.

在这个方向上 Simon 的一个重要贡献是 CaMeRa (computation with multiple representation,具有多种表达的计算)<sup>[16]</sup>.这个系统是他和 Hermina Tabachneck-Schilf 以及 M. Anthony Leonardo 于1997年合作开发的计算系统,用来模拟专家用多种表达方式(即图形(diagrammatic)表达与文字(verbal)表达)解决问题的过程.它是用 OPS5 (一种基于 Common LISP 的产生式系统语言)实现的.

CaMeRa 的一个重要特征是,不同的模态用不同的表达(数据结构及其操作).系统有一个图形短时记忆(pSTM)和图形长时记忆(pLTM)用于图形表达,有一个文字短时记忆(vSTM)和文字长时记忆(vLTM)用于文字表达.外界的图形表达于外在图形显示(一个位图 bitmap)中,用于内部图形表达的 pSTM,又叫做心理眼(mind's eye),由一个低层次的视觉缓冲区(vision buffer,也是一个位图,模拟主视觉皮层用)和一个由连接节点构成的高层次的空间和对象的信息结构所构成.一个小型的平行网络用于处理视觉缓冲区中的低层次图形信息.文字记忆结构模型是命题表.这两种表达的共同特点是这两种知识都是由小的组块(Chunk)组成,这些组块可以在同一模态内或者跨模态而关联连接.然而只能由他们所在模态的短时记忆来更动.基于规则的过程用于高水平的图形的以及文字的推理中.模型中没有目标堆栈.

CaMeRa 图形和文字表达是相互互补的表达.因而能得到两种表达特有的优势.例如:CaMeRa 成功地模拟了经济学专家是如何处理供求关系的.它模拟了专家用黑板上的图形来表达这种关系.如果物价高于了平衡点,CaMeRa 就会像专家一样察看供求关系图,发现有剩余(Surplus),然后进行文字推理,得出剩余将导致物价降低的结果.

CaMeRa 的另一个重要特征是心理图像(mental image)与视觉刺激非常相似.这样做是基于心理眼假说:由观看外在图画产生的心理图像于短时记忆中的表达与由对这个图画的记忆而形成的心理图像有相同的图像结构.在 CaMeRa 中,图形短时记忆中有视觉信息与文字信息的关联(而外部显示则

没有),因而能由存在内部记忆中的信息组成心理图像.

例如,CaMeRa 能过模拟测试者心理图像,作为物理过程的动态表达,来帮助他们理解 Einstein 的 1905 年相对论论文:论动体的电动力学[17]. 在仔细地定义了时钟的同时性,提出了相对性原理和光速不变原理以后,Einstein 邀请他的读者去想像光线沿着一个运动着的杆的运动的过程并由此推导出导致 Lorentz 变换的关键方程:“有一个给定的静态的刚性杆...,现在让我们想像这个杆的中轴线与静止坐标系的  $X$  坐标轴平行,然后,有一个速度为  $v$ ,方向沿着  $X$  轴增加的方向的匀速运动作用到杆上.....让我们进一步想象在杆的两端  $A$  和  $B$  各有一时钟,它们与静止坐标系的时钟同步,进一步想像一束光线在时间  $t_A$  由  $A$  端发出,在时间  $t_B$  时由  $B$  端反射回来,在于时间  $t_A$  回到  $A$  端,考虑到光速不变的原理,有

$$t_B - t_A = r_{AB} / (c - v), \quad (1)$$

及

$$t_A - t_B = r_{AB} / (c + v). \quad (2)$$

式中: $r_{AB}$  是运动的杆在静止坐标系中测量的长度. $c$  是光速. 请注意,按 Galilean 变换,式(1)和式(2)的右边应该相同,均等于  $r_{AB} / c$ . 如果不去“看”这个过程,一个人是很难(如果还有可能的话)推导出这两个方程来的. CaMeRa 能够模拟从按照阅读材料形成心理图像,检查心理图像,直到导出这些方程的全过程.

1) 整合阅读材料及记忆中的信息,形成可控的动态心理图像,以表达光沿着动杆运动的过程<sup>[18-19]</sup>,这个图像可以由图形短时记忆中视觉缓冲区中的位图模拟.

2) 观察心理图像所显示的过程,推出量之间的定性联系,例如,光从  $A$  到  $B$  的运行距离(由静止坐标系测量)大于杆长<sup>[20]</sup>;

当用 CaMeRa 模拟时,在图形短时记忆中的信息及存于文字短时记忆和文字长时记忆中的信息,均在此过程被调用.

3) 结合光速不变原理,推出两变量之间的定量关系.如,光从  $A$  到  $B$  的过程<sup>[20]</sup>,CaMeRa 能够用文字推理结合观察心理图像模拟:

a) 光从  $A$  到  $B$  运行距离大于杆长(定性关系,由观察心理图像获得).

b) 光从  $A$  到  $B$  运行距离大于杆长  $\Rightarrow$

光运行距离 =  $r_{AB}$  + 某种东西(半定性关系)

c) 某种东西 = 光运行时间  $\times$  杆速(半定性关系,由观察心理图像获得).

d) 某种东西 = 光运行时间  $\times$  杆速  $\Rightarrow$

某种东西 =  $(t_B - t_A) \times V$  (半定量关系).

e) 某种东西 =  $(t_B - t_A) \times V \Rightarrow$

光运行距离 =  $r_{AB} + (t_B - t_A) \times V$  (半定量关系).

f) 光运行距离 =  $(t_B - t_A) \times C$  (光速不变原理).

g) 结合 e), f)  $\Rightarrow$

$(t_B - t_A) \times C = r_{AB} + (t_B - t_A) \times V$  (定量关系).

h)  $(t_B - t_A) \times C = r_{AB} + (t_B - t_A) \times V \Rightarrow$

$$t_B - t_A = r_{AB} / (C - v) \quad (3)$$

### 3 结束语

Simon 离开我们已经五年了,他的科研方法,如上面两节所示,由检查人的行为(例如,通过口头报告分析)到建立计算模拟模型,再通过检查计算模型来加深对人的行为的理解,依然值得我们效仿.

1997 年,在他离去的前 4 年,他为他的传记《我生活的种种模式》的中文版写了中译本序.这个序已经全文翻译成中文,登在文献[21]之中,英文从未发表.下面摘出其中部分英文原文,在 AI 诞生 50 周年纪念时,与 AI 的探索者共享.

“As I have had the good fortune to live through the years of the birth of the modern electronic computer and the field of artificial intelligence to which it gave rise, a large part of my story is the story of those exciting years. My chief hope for this autobiography is that it may give young people considering a career in science, or just entering such a career, some picture of the excitement of life in science. Of course the picture it paints refers to years, many of which are long past, and mostly to a land that is very far from China; but the urge that a scientist feels to explore the unknown is not peculiar to any time or to any special part of the globe. In whatever century and whatever land we spend our lives, we can respond to that urge and experience the satisfactions of finding new ideas and new things of value to humanity.”

## 参考文献:

- [1] SIMON H. Models of my life[M]. New York: Basic-Books, 1991.
- [2] NEWELL A, SHAW J, SIMON H. Empirical explanations of the logic theory machine: A case study in heuristic[A]. Proceedings of the Western Joint Computer Conference[C]. New York: McGraw - Hill, 1963.
- [3] NEWELL A, SHAW J, SIMON H. The process of creative thinking[J]. Models of Thought, 1979(1): 144 - 174.
- [4] NEWELL A, SIMON H. Human problem solving[M]. Englewood Cliffs, NJ: Prentice-Hall, 1972.
- [5] NEWELL A. Unified theories of cognition[M]. Cambridge, MA: Harvard University Press, 1990.
- [6] NEVES D. A computer program that learns algebraic procedures by examining examples and working problems in a textbook[A]. In: Proceedings of the Second National Conference of the Canadian Society for Computational Studies of Intelligence[C]. [S.L.], 1978.
- [7] ZHU X, SIMON H. Learning mathematics from examples and by doing[J]. Cognition and Instruction, 1987, 4(3): 137 - 166.
- [8] DEJONG G, MOONEY R. Explanation-based learning: An alternative view[J]. Machine Learning, 1986, 1(2): 145 - 176.
- [9] MITCHELL T, KELLER R, KEDARCABELLI S. Explanation-based generalization: a unifying view[J]. Machine Learning, 1986, 1(1): 47 - 80.
- [10] MITCHELL T, ALLEN J, CHALASANI P, et al. Theo: a framework for self-improving system[M]. NJ: Lawrence Erlbaum Associate, 1989.
- [11] QIN Y, MITCHELL T, SIMON H. Using EBG to simulate human learning from examples and learning by doing[A]. In: Proceedings of the fifth Artificial Intelligence Research Symposium[C]. [S.L.], Florida, 1992.
- [12] ZHU X, LEE Y, SIMON H, ZHU Dan. Cue recognition and cue elaboration in learning from examples[A]. In: Proceedings of National Academy of Sciences of USA[C]. [S.L.], 1996.
- [13] ZHU X, ZHU D, LEE Y, SIMON H. Cognitive theory to guide curriculum design for learning from examples and by doing[J]. Journal of Computers in Mathematics and Science Teaching, 2003, 22(4): 285 - 322.
- [14] LARKIN J, SIMON H. Why a diagram is (sometimes) worth ten thousand words[J]. Cognitive Science, 1987(11): 65 - 99.
- [15] SIMON H. Diagrammatic reasoning, cognitive and computational perspectives[M]. Cambridge: The MIT Press, 1995.
- [16] TABACHNECK H S, LEONARDO A, SIMON H. CaMeRa: a computational model of multiple representations[J]. Cognitive Science, 1997, 21(3): 305 - 350.
- [17] EINSTEIN A. On the electrodynamics of moving bodies, 1905[M]. Dover Publications, 1923.
- [18] QIN Y, SIMON H. Image as process representation in problem solving[A]. In: Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society[C]. Hillsdale, NJ: Lawrence Erlbaum Associate, 1992.
- [19] QIN Y, SIMON H. Diagrammatic reasoning, cognitive and computational perspectives[M]. Cambridge: The MIT Press, 1995.
- [20] QIN Y. From language to mental images to equations[D]. Pennsylvania, Pittsburgh: Carnegie Mellon University, 1992.
- [21] SIMON H. 我生活的种种模式[M]. 上海: 东方出版社, 1998.

## 作者简介:



秦裕林, 1947年生. 浙大心理系兼任教授; 美国卡内基·梅隆大学(CMU)高级研究心理学家. 北航计算机硕士, CMU 认知心理学博士, 曾与导师 Herbert A. Simon 共同署名发论文 6 篇, 与 Anderson(美科学院院士)共同署名发论文 9 篇, 其中 2 篇在美科学院院刊发表. 目前研究方向为计算认知体系结构(Computational Cognitive Architecture) ACT - R 及其应用. E-mail: yulingq@yahoo.com