



21世纪高校特色教材

人工智能 与机器翻译

主 编 杨宪泽



RENGONG ZHINENG
YU
JIQUEFANYI



西南交通大学出版社
<http://press.swjtu.edu.cn>

TP18
124

人工智能与机器翻译

主编 杨宪泽

西南交通大学出版社

·成都·

内 容 简 介

本书的内容分为两大部分：一是人工智能原理（知识表示、知识推理、知识学习中的主流技术及其原理）；二是机器翻译原理与应用（按照一般机器翻译系统应该具备的各个处理步骤的顺序进行讲述，实验时通过在每个步骤设计一个小的软件来逐步完成整个翻译软件的设计与实现）。此外，为使学生巩固所学的人工智能原理和机器翻译原理知识，本书还组织了单词与词组的处理与分析、词义消歧、句法（语法）与语法分析和机器学习方面的内容。

本书的许多内容取材于编著者们的科研积累，深入浅出，内容详实。

本书可作为计算机科学与技术、软件工程、信息安全等专业的本科教材，也可以作为相关方向的研究生教材，同时也可以供从事这方面研究的科技人员参考。

图书在版编目 (C I P) 数据

人工智能与机器翻译 / 杨宪泽主编. —成都: 西南交通大学出版社, 2006.2
ISBN 7-81104-208-8

I. 人... II. 杨... III. ①人工智能—高等学校—教材②机器翻译—高等学校—教材 IV. ①TP18②H085

中国版本图书馆 CIP 数据核字 (2006) 第 004449 号

人工智能与机器翻译

主编 杨宪泽

*

责任编辑 张宝华

特邀编辑 孙康江

责任校对 韩松云

封面设计 支鑫淼

西南交通大学出版社出版发行

(成都二环路北一段 111 号 邮政编码: 610031 发行部电话: 028-87600564)

<http://press.swjtu.edu.cn>

E-mail: cbsxx@swjtu.edu.cn

四川煤田地质制图印刷厂印刷

*

成品尺寸: 185 mm×240 mm 印张: 16.5

字数: 422 千字

2006 年 2 月第 1 版 2006 年 2 月第 1 次印刷

ISBN 7-81104-208-8/H·016

定价: 36.00 元

图书如有印装问题 本社负责退换

版权所有 盗版必究 举报电话: 028-87600562

**本书可作为计算机科学与技术、软件工程、信息安全等
本科专业的专业基础教材**

主 编：杨宪泽

副主编：谈文蓉

编 委：刘玉萍 刘明志 朱 晖 肖 明 李建阳

前 言

人工智能学科 20 世纪 80 年代迅速发展，这是人工智能研究的第二次高潮。主要原因是当时日本人研究“第五代智能机”的呼声如火如荼，世界各国纷纷响应，中国也随之响应并称为“迎接新技术革命挑战”。这一阶段，我国不仅在计算机智能技术研究方面投入比较大的人力、物力，而且微机应用也开始普及，高校也开办了许多计算机专业。当时，人工智能原理成为计算机科学技术主干课程之一，也是重中之重的课程。

人工智能技术涉及的领域非常广，技术非常复杂。对于计算机本科生来说，人工智能原理非常难学。针对这一情况，各个学校的教材版本非常多，但主要有以下三类：

- (1) 几乎涉及人工智能领域全部技术的教材；
- (2) 以专家系统为主线讲授人工智能技术的教材；
- (3) 从知识工程角度讲授人工智能技术的教材。

第 (1) 类教材虽然对人工智能技术面面俱到，也许在当时的历史背景下应该这样，但本科生难以接受，效果差。另外，学生实验不好设计，当时的 LISP 等人工智能语言的实验流于形式。

第 (2) 类教材迄今看来仍然是一种好的版本，但专家系统在人工智能技术中相对比较简单，而且学生实验必须花费很多的时间，最后许多学校不能开实验，这又成了一个问題。

第 (3) 类教材实际上只讲授人工智能涉及到的软件技术，但也存在怎样安排学生实验的问题，最后没有得到众多高校的认同。

进入 20 世纪 90 年代，随着日本人研究“第五代智能机”的失败，加之智能技术在许多方面并没有取得预期的成功，而计算机科学技术在其他方面发展非常迅速（如网络技术、计算机信息处理等），人工智能技术的研究相比 80 年代初期开始降温。因此，各个高校计算机学科专业对人工智能原理的重视程度大不如 80 年代，该课程大多选择第 (1) 类教材，而学生实验几乎仅仅是一种形式。

而今，许多从事人工智能技术研究的学者和教育工作者都认识到：人工智能仍然是一门不断发展、非常有潜力的学科，它在计算机科学中的重要地位是不可动摇的。人工智能方面不管是以前、现在，还是将来，它所出现的成果都是人类认识模糊或者难以解决的问题。因此，应该重视人工智能技术的本科教育。

基于上述认识，我们应该设计、选择、编写一本更适宜计算机专业本科生的人工智能方面的教材。这本教材在人工智能基础理论方面重点突出，内容又不太多，而且容易理解消化；在实验方面，花费时间合理，能够巩固人工智能基础理论方面的知识，因此，既有意义，又能使学生感到兴趣。几年来的课程实践表明：将机器翻译技术与人工智能原理结合到一门课程中进行讲授，是一次较为成功的探索。

本书取材于编著者们多年的科研积累，曾先后得到国家民委重点科研项目的资助，四川省科技攻关项目的资助，西南民族大学重点科研项目的资助，西南民族大学精品课程建设项目的资助，这些资助使得编著者们获得了一定量的成果，为本教材的完成奠定了较好的基础。

本书第一章由杨宪泽、李建阳老师编著；第二、三章由杨宪泽、朱晖老师编著；第四章由杨宪泽、刘明志老师编著；第五章由杨宪泽、谈文蓉老师编著；第六、七章由杨宪泽、肖明老师编著。全书由杨宪泽教授统稿。

本书还配有相应的教学辅助课件，该课件由杨宪泽、谈文蓉、刘玉萍、刘明志、朱晖、肖明、李建阳老师共同制作，需要者可与作者联系。

由于人工智能技术的新颖性和复杂性，加之编著者的水平所限，书中难免有错误和不妥之处，希望读者谅解和及时批评指正。

编著者

2005年10月于成都

目 录

第1章 总 论	1
1.1 人工智能	2
1.2 对人工智能的新认识	6
1.3 人工智能的研究与应用领域	11
1.4 智能软件示例	20
1.5 机器翻译	22
习题一	31
第2章 相关知识表示方法	32
2.1 知识表示概述	33
2.2 一阶谓词逻辑表示方法	34
2.3 单元表示方法	37
2.4 与 / 或图表示方法	41
2.5 语义网络表示方法	42
2.6 知识表示的小结	55
习题二	56
第3章 产生式系统及其搜索方法	57
3.1 产生式系统	58
3.2 产生式系统应用举例	63
3.3 产生式系统的搜索(控制)策略	65
3.4 图搜索算法	67
3.5 产生式系统的规则问题	77
3.6 产生式系统的不确定性问题	79
3.7 产生式系统的常用开发语言简介	83
3.8 产生式系统设计技巧	84
习题三	85
第4章 机器翻译方法	87
4.1 机器翻译的四种实现方法	88
4.2 基于实例的机器翻译改进方法	93

4.3 基于实例的机器翻译方法实践	95
4.4 与机器翻译相似的自然语言检索接口	102
4.5 机器翻译评价	104
4.6 机器翻译的优劣 —— 对机器翻译系统的评价	106
习题四	112
第5章 单词及词组的处理与分析	113
5.1 机器词典概述	114
5.2 自动分词	115
5.3 交集型和多义组合型歧义切分	117
5.4 词语的排序、检索、词库	119
5.5 词语的分类与兼类问题	119
5.6 词语处理的其他问题	123
5.7 词义消歧	129
习题五	150
第6章 句法(语法)与语义理论及分析	151
6.1 句法分析与用作翻译的理论基础	152
6.2 语法分析在机器翻译中的应用	157
6.3 其他语法理论介绍	170
6.4 汉语和英语语法简介	184
6.5 语义分析	190
习题六	202
第7章 相关机器学习	203
7.1 机器学习概述	204
7.2 从例子中学习	208
7.3 类比学习	217
7.4 解释学习	221
7.5 遗传式学习	226
7.6 统计学习	230
7.7 机器学习在机器翻译中的应用	240
习题七	247
附 录	249
参考文献	254

第 1 章 总 论

人工智能是计算机科学的一个分支，在近五十年的发展过程中，人工智能学科经历了不少的争论、困难和挑战，但是它还是不断向前，迅速发展，已经成为当今社会主要的代表技术之一，有着举足轻重的地位和影响。人工智能的任务旨在模拟人类行为和认知过程，但人类的行为是个非常复杂的过程，至今仍未能被完全解释。因此，人工智能涉及的技术非常多也异常复杂，这给我们的学习带来困难。基于这一点，本章将概述人工智能的起源与发展、人类智能的计算机模拟、人工智能的研究与应用领域，最后概述本书学习人工智能的主线——机器翻译技术。

1.1 人工智能

1.1.1 人工智能的起源与发展

对于人工智能的发展来说,在20世纪30年代和40年代,科学界发现了两件最重要的事:数理逻辑(它从19世纪末起就获得迅速发展)和关于计算的新思想。弗雷治(Frege)、怀特赫德(Whitehead)、罗素(Russell)和塔斯基(Tarski)以及另外一些人的研究表明,推理的某些方面可以用比较简单的结构加以形式化。数理逻辑后来成为人工智能研究的一个活跃领域,其部分原因是由于一些逻辑演绎系统可以在计算机上实现。不过,即使在计算机出现之前,逻辑推理的数学公式也为人们建立了计算与智能关系的概念。

丘奇(Church)、图灵(Turing)和其他一些人关于计算本质的思想,指出了形式推理与即将发明的计算机之间的联系。在这方面的一个重要工作是关于计算和符号处理的理论。第一批数字计算机(实际上为数字计算器)看来不包含任何真实智能。早在这些机器设计之前,丘奇和图灵就已发现,数字并不是计算的主要方面,它们仅仅是一种解释机器内部状态的方法。被称为人工智能之父的图灵,不仅创造了一个简单的通用的非数字计算模型,而且直接证明了计算机可能以某种被理解为智能的方法工作。

1956年夏季,年轻的美国学者麦卡锡(McCarthy)、明斯基(Minsky)、朗彻斯特(Lochester)和香侬(Shannon)共同发起,邀请莫尔(More)、塞缪尔(Samuel)、纽厄尔(Newell)和西蒙(Simon)等参加在美国的达特茅斯(Dartmouth)大学举办了一次长达两个月的研讨会,认真热烈地讨论用机器模拟人类智能的问题。会上,首次使用了人工智能这一术语。这是人类历史上第一次人工智能研讨会,标志着人工智能学科的诞生,具有十分重要的历史意义。这些从事数学、心理学、信息论、计算机和神经学研究的年轻学者,绝大多数都成为著名的人工智能专家,50年来为人工智能的发展作出了重要贡献。

1969年召开了第一届国际人工智能联合会议(International Joint Conference on AI),此后每两年召开一次;1970年《人工智能》国际杂志(International Journal of AI)创刊。这些对开展人工智能国际学术活动,促进人工智能的研究和发展起到积极作用。

值得一提的是控制论思想对人工智能早期研究的影响,正如艾伦·纽厄尔(Allen Newell)和赫伯特·西蒙(Herbert Simon)1972年在他们的优秀著作《人类问题求解》(Human Problem Solving)的“历史补篇”中指出的那样,20世纪中叶人工智能的奠基者在人工智能研究中出现了几股强有力的思潮。维纳(Wiener)、麦克洛克(McCulloch)和其他一些人提出的控制论和自组织系统集中地讨论了“局部简单”系统的宏观特性。我国优秀的科学家钱学森提出的“工程控制论”开辟了控制论的新分支,是对控制论的重大贡献。控制论影响了许多领域,因为控制论的概念跨接了许多领域,把神经系统的工作原理与信息理论、控制理论、逻辑以及计算联系起来。控制论的这些思想是时代思潮的一部分,而且在许多情况下影响了许多早期人工智能工作者,成为他们的指导思想。



最终把这些不同思想连接起来的是由巴贝奇 (Babbage)、图灵、冯·诺依曼 (Von Neumann) 和其他一些人所研制的计算机。在机器的应用成为可能之后,人们就开始试图编写程序以解决智力测验、下棋以及把文本从一种语言翻译成另一种语言等问题。这是第一批人工智能程序。对于计算机来说,促使人工智能发展的是什么?出现在早期设计中的许多与人工智能有关的计算概念,包括存储器和处理器的概念、系统和控制的概念,以及语言的程序级别的概念。不过,引起新学科出现的新机器的唯一特征是这些机器的复杂性,它促进了对描述复杂过程方法的新的更直接的研究(采用复杂的数据结构和具有数以百计的不同步骤的过程来描述这些方法)。

几十年来,人工智能的应用研究取得明显进展。首先,专家系统 (expert system) 显示出强大的生命力。被誉为“专家系统和知识工程之父”的费根鲍姆 (Feigenbaum) 所领导的研究小组于 1968 年研究成功第一个专家系统 DENDRAL,用于质谱仪分析有机化合物的分子结构。1972—1976 年,费根鲍姆小组又开发成功 MYCIN 医疗专家系统,用于抗生素药物治疗。此后,许多著名的专家系统,如 PROSPECTOR 地质勘探专家系统、CASNET 青光眼诊断治疗专家系统、RI 计算机结构设计专家系统、MACSYMA 符号积分与定理证明专家系统,ELAS 钻井数据分析专家系统和 ACE 电话电缆维护专家系统等被相继开发,为工矿数据分析处理、医疗诊断、计算机设计、符号运算和定理证明等提供强有力的工具。1977 年,费根鲍姆进一步提出了知识工程 (knowledge engineering) 的概念。20 世纪 80 年代,专家系统和知识工程在全世界得到迅速发展。在开发专家系统过程中,许多研究者获得共识,即人工智能系统是一个知识处理系统,而知识表示、知识利用和知识获取则成为人工智能系统的三个基本问题。

近年来,机器学习和神经网络研究深入开展,形成高潮。同时,不同人工智能学派间的争论也非常热烈。这些都推动人工智能研究的进一步发展。

我国的人工智能研究起步较晚。纳入国家计划的研究 (“智能模拟”) 始于 1978 年; 1984 年召开了智能计算机及其系统的全国学术讨论会; 1986 年起把智能计算机系统、智能机器人和智能信息处理 (含模式识别) 等重大项目列入国家高技术研究计划; 1993 年起,把智能控制和智能自动化等项目列入国家科技攀登计划。1981 年起,相继成立了中国人工智能学会 (CAAI)、全国高校人工智能研究会、中国计算机学会人工智能与模式识别专业委员会、中国自动化学会模式识别与机器智能专业委员会、中国软件行业协会人工智能协会、中国智能机器人专业委员会、中国计算机视觉与智能控制专业委员会以及中国智能自动化专业委员会等学术团体。1989 年首次召开的中国人工智能控制联合会议 (CJCAI) 至今已召开多次。已有许多国内自编的人工智能专著和教材公开出版,《模式识别与人工智能》杂志已于 1987 年创刊。现在,我国已有数以万计的科技人员和大学师生从事不同层次的人工智能研究与学习,人工智能研究已在我国深入开展,它必将为促进其他学科的发展和我国的现代化建设作出新的重大的贡献。

1.1.2 人类智能的计算机模拟

人类的认知过程是个非常复杂的行为,至今仍未能被完全解释。人们从不同的角度对它进行研究,从而形成诸如认知生理学、认知心理学和认知工程学等相关学科。对这些学科的深入研究已超出本书范围,作为教材,这里仅讨论几个与人工智能有密切关系的问题。

1. 研究认知过程的任务

人的心理活动具有不同的层次，它可与计算机的层次相比较。心理活动的最高层级是思维策略，中间一层是初级信息处理，最低层级为生理过程，即中枢神经系统、神经元和大脑的活动。与此相应的是计算机的程序、语言和硬件，如图 1-1 所示。

研究认知过程的主要任务是探求高层次思维决策与初级信息处理的关系，并用计算机程序来模拟人的思维策略水平，而用计算机语言模拟人的初级处理过程。

令 T 表示时间变量， x 表示认知操作 (cognitive operation)， x 的变化 Δx 为当时机体状态 S (机体的生理和心理状态以及脑子里的记忆等) 和外界刺激 R 的函数。当外界刺激作用到处于某一特定状态的机体时，机体便发生变化，同样，计算机也以类似的原理进行工作。在规定时间内，计算机存储的记忆相当于机体的状态；计算机的输入相当于机体施加的某种刺激。在得到输入后，计算机便进行操作，使得其内部状态随时间发生变化。可以从不同的层次研究这种计算机系统。这种系统以人的思维方式为模型进行智能信息处理 (intelligent information processing)。显然，这是一种智能计算机系统。人们设计用于特定领域的这种高水平智能信息处理系统 (也称为专家系统) 是研究认知过程的一个具体而又重要的目标。例如，一个具有智能信息处理能力的自动控制系统就是一个智能控制系统，它可以是专家控制系统，或者是智能决策系统等。

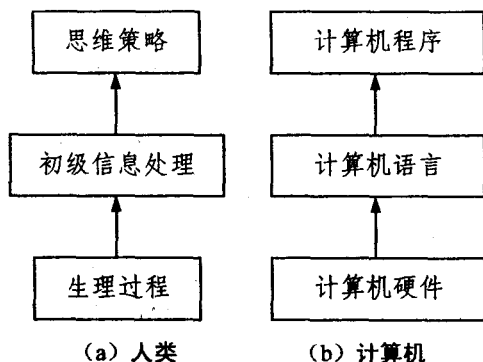


图 1.1 人类认知活动与计算机的比较

在规定时间范围内，计算机存储的记忆相当于机体的状态；计算机的输入相当于机体施加的某种刺激。在得到输入后，计算机便进行操作，使得其内部状态随时间发生变化。可以从不同的层次研究这种计算机系统。这种系统以人的思维方式为模型进行智能信息处理 (intelligent information processing)。显然，这是一种智能计算机系统。人们设计用于特定领域的这种高水平智能信息处理系统 (也称为专家系统) 是研究认知过程的一个具体而又重要的目标。例如，一个具有智能信息处理能力的自动控制系统就是一个智能控制系统，它可以是专家控制系统，或者是智能决策系统等。

2. 智能信息处理系统的假设

可以把人看成一个智能信息处理系统。

信息处理系统又叫符号操作系统 (symbol operation system) 或物理符号系统 (physical symbol system)。所谓符号就是模式 (pattern)。任一模式，只要它能与其他模式相区别，它就是一个符号。不同的汉语拼音字母或英文字母就是不同的符号。对符号进行操作就是对符号进行比较，从中找出相同的和不同的符号。物理符号系统的基本任务和功能就是辨认相同的符号和区别不同的符号。为此，这种系统就必须能够辨别出不同符号之间的实质差别。符号既可以是物理符号，也可以是头脑中的抽象符号，或者是电子计算机中的电子运动模式，还可以是头脑中神经元的某些运动方式。一个完善的符号系统应具有下列 6 种基本功能：

- (1) 输入符号 (input);
- (2) 输出符号 (output);
- (3) 存储符号 (store);
- (4) 复制符号 (copy);
- (5) 建立符号结构：通过找出各符号间的关系，在符号系统中形成符号结构。



(6) 条件性迁移 (conditional transfer): 根据已有符号, 继续完成活动过程。

如果一个物理符号系统具有上述全部 6 种功能, 并能够完成这个全过程, 那么它就是一个完整的物理符号系统。人能够输入信号, 如用眼睛看, 用耳朵听, 用手触摸等。计算机也能通过卡片或纸带打孔、磁带或键盘打字等方式输入符号。人具有上述 6 种功能; 现代计算机也具备物理符号系统的这 6 种功能。

假设 任何一个系统, 如果它能表现出智能, 那么它就必定能够执行上述 6 种功能。反之, 任何系统如果具有这 6 种功能, 那么它就能够表现出智能, 这种智能指的是人类所具有的那种智能。把这个假设称为物理符号系统的假设。

物理符号系统的假设伴随有 3 个推论, 或称为附带条件。

推论一 既然人具有智能, 那么他(她)就一定是一个物理符号系统。人之所以能够表现出智能, 就是基于他的信息处理过程。

推论二 既然计算机是一个物理符号系统, 那么它就一定能够表现出智能。这是人工智能的基本条件。

推论三 既然人是一个物理符号系统, 计算机也是一个物理符号系统, 那么我们就能够用计算机来模拟人的活动。

值得指出, 推论三并不一定是从推论一和推论二推导出来的必然结果。因为人是物理符号系统, 具有智能; 计算机也是一个物理符号系统, 也具有智能, 但它们可以用不同的原理和方式进行活动, 所以, 计算机并不一直都是模拟人的活动的, 它可以编制出一些复杂的程序来求解方程式, 进行复杂的计算, 计算机的这种运算过程未必就是人类的思维过程。

按照人类的思维过程来编制计算机程序, 这项工作就是人工智能的研究内容, 也是智能控制的主要研究内容。如果做到了这一点, 那么就可以用计算机在形式上来描述人的思维活动过程, 或者建立一个理论来说明人的智力活动过程。

3. 人类智能的计算机模拟

帕梅拉·麦考达克(Pamela Mccorduck)在她的著名的人工智能历史研究《机器思维》(Machine Who Think, 1979)中曾经指出: 在复杂的机械装置与智能之间存在着长期的联系。从几世纪前出现的神话般的复杂巨钟和机械自动机开始, 人们已对机器操作的复杂性与自身的智能活动进行直接联系。今天, 新技术已使人类所建造的机器的复杂性大为提高, 现代电子计算机要比以往的任何机器复杂几十倍、几百倍, 甚至几千倍以上。

计算机的早期工作主要集中在数值计算方面。然而, 人类最主要的智力活动并不是数值计算, 而是在逻辑推理方面。物理符号系统假设的推论一也说明, 人有智能, 所以他是一个物理符号系统; 推论三指出, 可以编写出计算机程序去模拟人类的思维活动。这就是说, 人和计算机这两个物理符号系统所使用的物理符号是相同的, 因而计算机可以模拟人类的智能活动过程。计算机的确能够很好地执行许多智能功能, 如下棋、证明定理、翻译语言文字和解决难题等。这些任务是通过编写与执行模拟人类智能的计算机程序来完成的。当然, 这些程序只能接近于人的行为, 而不可能与人的行为完全相同。此外, 这些程序所能模拟的智能问题, 其水平还是很有限的。

作为例子, 可以考虑下棋的计算机程序。现有的国际象棋程序是十分熟练的、具有人类专家



棋水平的最好实验系统，但是下得没有像人类国际象棋大师那样好。该计算机程序对每个可能的走步空间进行搜索，它能够同时搜索几千种走步。进行有效搜索的技术是人工智能的核心思想之一。不过，计算机不能战胜最好的人类棋手，其原因在于：向前看并不是下棋所必须具有的一切，需要彻底搜索的走步又太多，在寻找和估计替换走步时并不能确信能够导致博弈的胜利。国际象棋大师们具有尚不能解释的能力。一些心理学家指出，当象棋大师们盯着一个棋位时，在他们的脑子里出现了几千盘重要的棋局，这大概能够帮助他们决定最好的走步。至今还未能设计出辨别这种棋局的计算机程序。

近年来，智能计算机的研究取得许多重大进展。对神经型智能计算机的研究就是一个新的范例，它必将为模拟人类智能作出新的贡献。

神经计算机(neural computer)能够以类似人类的方式进行“思考”，它力图重建人脑的形象。据日本通产省(MITI)报道，对神经计算机系统的可行性研究早于1989年4月底完成，并提出了该系统的长期研究计划的细节。在美国、英国、中国和其他一些国家，都有众多的研究小组投入对“神经网络”的研究，一个“神经网络热”正在形成。据预测，神经计算机将在不久的将来进入实用阶段，并将有产品投放市场。

人脑这个神奇的器官能够复制大量的交互作用，快速处理极其大量的信息，同时执行几项任务。迄今为止的所有计算机，基本上都未能摆脱冯·诺依曼机的结构，只能依次对单个问题进行“求解”。即使是20世纪80年代初期的并行处理计算机，其运行性能仍然十分有限。人们期望，对神经计算(neural computing)的研究将造出神经计算机，大大提高信息处理能力，达到更高的人工智能水平。

1.2 对人工智能的新认识

1.2.1 再述人工智能

一般认为，人工智能是计算机科学的一个分支，旨在模拟人类行为和认知过程。而自然语言处理和理解又是人工智能的一个分支，机器翻译又是自然语言理解和处理的一个分支。鉴于这种关系，在机器翻译理论和应用研究中肯定会使用许多人工智能技术，所以有必要在总论这一章中再加入对人工智能情况的新的认识。

三十多年来，人工智能在自然语言理解、机器翻译、模式识别、计算机视觉、定理证明、专家系统、机器学习和机器人学等方面取得了成绩，引人注目，因此，如何在机器翻译的理论和应用研究中引进智能技术以提高系统整体性能，是一项重要的任务。

计算机和人脑的物质结构大不一样，一个是无生命的机器，一个是由一千亿神经元组成的活生生的活体。但是，计算机软件所表现出的功能和人的认知过程却是类同的，即两者的工作原理是一致的，都是信息加工系统：即输入信息、进行编码、存储记忆、作出决策和输出结果。由于计算机和人的认知过程在信息加工原理上一致，因此可以把计算机当成实验工具，模拟人的认知活动。上述认识有一个过程，计算机能否智能，在计算机发展的初期就提出来了。图灵在他1950



年发表的一篇论文“计算机与智能”中，回答了这个问题。在这篇论文中，他设计了一个试验，后来成为著名的图灵试验。此实验说明，人们可以通过编制程序使计算机体现出某种智能。几乎在同时，香农也指出了与计算机下象棋的可能性，后来得以实现。

人工智能技术的发展经历了曲折的过程。开始，研究者们认为：人工智能作为一门科学，也应该像数学，物理学等学科那样能够发现一些普遍的定律，把这些定律同强有力的计算机结合起来应用到各个领域就可以产生不可估价的成效。因此，许多研究者致力于人类思维普遍规律的探索。人工智能进入了推理方法和问题求解模型的一般性研究。这一阶段，的确发现了一些通用问题求解技术和各种搜索策略。然而，现实世界是复杂的，问题是多种多样的。随着研究过程的经验积累，研究者们逐渐认识到：人们在探索自然界的长期实践中所形成的解决问题的方法和手段，因事甚至因人而变化，所以，试图找出一种或几种通用的问题表示方法和处理过程去描述人类一切智能行为的想法，不说根本不可能，至少也是不现实的。一些通用性的问题求解策略其通用性也是有限的，对于求解很复杂的问题最终还显得不够，需要特定领域内所特有的技巧。

可以认为，智能技术能取得今天的成绩，是一个从特殊到一般，又从一般到特殊相辅相成的过程，即不断从模拟特殊问题求解的智能行为开始，然后再从中抽取一般规律性问题，以使研究推向深入。然而，由于人的智能是多种能力的综合，在处理不同问题时，又显示出各自独有的特性，所以设计具体问题时又必须从一般到特殊。

今天，智能技术已无孔不入，在计算机应用中发挥着巨大的作用，研究的热潮一浪高过一浪。但对此必须保持清醒的认识，因为即使是已能实用并引起很多行业兴趣的专家系统，若认真从人工智能角度看，其中的技术也并不很高：一些自然语言理解系统和语音识别系统，经历数年攻关，目前仍没有原则性突破。此外，计算机并不等于人脑，人类至今不仅对哺乳动物、鱼类、甚至对与人类进化过程相差数亿年的最低级动物（如水螅）的神经机制都搞不清楚，要搞清人脑机制为时遥远。然而这并不是否认智能技术。虽然我们不能造出像人一样思维的计算机，但却有可能用计算机模拟人的某些思维过程，解决一些复杂的问题。例如，虽然对鸟类飞行的神经机制至今也没有搞清楚，但人类根据鸟飞行的空气动力学原理制造的飞机却远远超出了鸟类飞行的能力。类似地，虽然我们不能精确地了解人类思维的神经系统活动机制，但根据人类解决问题的外部启发式特征设计的系统却可以在一定范围内和一定程度上解决只有人类才能解决的问题。在机器翻译系统中使用智能技术的原则是：设计机器翻译某一模块时，引进智能技术必须明显地提高模块性能，否则，就没有必要使用智能技术。若画蛇添足，只能降低翻译的速度，延长研制周期。

顺便指出，当前，研究神经网络是另一个支撑点。如何看待这一状况呢？当然，研究神经网络的主要目的不在于仿造一个人脑。人们已从鸟类的飞翔中获得启示，并与物理学相结合，创立了空气动力学并作为飞机设计的理论基础，同样，人们要从大脑神经网络的功能中获得启示，并与计算机科学相结合，创建新的人工智能理论，为设计面向智能信息处理的新一代计算机系统奠定坚实的理论基础。

1.2.2 通常软件系统和智能软件系统

一个智能软件系统也是一个计算机程序，因此，它同通常软件系统在程序的设计和实现方面有许多共同的特征。如一个智能系统的开发，也需要进行相应的分析、设计、编写、调试、维护



等阶段,各个阶段也会用到与通常程序开发相同的一些技术和手段;但另一方面,智能软件系统又是一个智能程序,它能体现出与通常程序所不同的一些特色和通常程序所没有达到的许多优良特性。

众所周知,计算机早期应用主要集中在科学计算和日常事务性工作数据处理。为使计算机实现一个确定范围的问题求解,程序设计人员需要对问题进行深入细致的研究,设计出一种逐步递进的求解步骤,称为算法。然后把这种算法的每一步骤用特定的程序设计语言翻译成程序输入计算机。计算机再按照程序所描述的步骤针对具体问题输入数据进行操作、求解。这种完全基于确定的算法的程序设计方法称之为通常方式的程序设计;相对来说这类算法具有三个性质:

(1) 通用性。算法应能求解问题范围内的全部问题,而不是只能解决其中的某些特殊问题。

(2) 确定性。算法中的问题求解状态、求解步骤应该是精确的、唯一的,并且可以被机械执行。

(3) 有效性。问题范围内的任何具体问题代入算法后,都可以经过有限步骤达到所期望的结果。

然而,由于现实世界的复杂性,一方面,还存在着大量的人类尚未充分认识的问题,程序设计人员目前尚不可能找到有效的算法;另一方面,有许多类型的问题,经过了严谨证明,不存在求解这类问题的算法;还有,对某些类型的问题,即使存在算法,也不可能付诸实现,因为一个算法的执行要受到时间和空间复杂性的限制,然而,这类问题的解需要几百小时、几千小时、甚至几千年以上才能解决。与通常程序设计相比,智能程序不是单纯的依靠算法,它结合了算法和启发方式。什么是启发方式?笔者认为,它是一些简化问题的法则,经验公式。从某种角度看,它仍然可能是算法,但却是特殊的算法。使用启发方式往往具有试探性,若成功,则说明问题的解过程进入捷径,但这些解还需人们判断其合理程度;若不成功,则要更换启发方式;如果更换启发方式仍失败,则说明法则不够或者问题无解。

与通常程序设计相比,智能程序也有以下三个性质:

(1) 局部性。智能程序可能仅适用于求解一类问题中那些被认为合理的或者是常见的问题。

(2) 试探性。通常程序要求问题的求解步骤是精确的、唯一的,不允许出错;而智能程序常采用一般情况下能保证正常工作的方法进行问题求解,当这个方法失败时,允许改用其他方法。

(3) 针对性。智能程序常利用求解问题的一些特殊规律,但这些特殊规律是针对性的,甚至是不精确的,它们经不起或者未经过严谨的理论证明,不能保证对每个具体问题有精确解或最优解。

以上从程序设计的角度讨论了智能程序和通常程序之间的差异,下面从研制智能软件系统和通常软件系统的角度再看一看两者的差异:

(1) 一般情况下,智能软件系统把求解问题的知识分为三级层次加以组织:数据级、知识库级和控制级。数据级知识是特定的具体问题所具有的,它们的组织像通常软件系统一样,不过智能软件系统中数据库结构可能更复杂;知识库级知识是法则和经验公式等的表达,可在不影响系统结构的前提下进行扩充完善;控制级知识表达了求解策略和推理模式。而在通常软件系统中,只有数据级和程序级两级结构。智能软件系统把知识库和推理机制分开使组织有较大的灵活性,便于知识的扩充、完善。

(2) 通常软件系统以数据为处理对象,而智能软件系统更强调符号处理。这是所有智能软件



系统与通常软件系统区别的一个重要特征。

(3) 智能软件系统可在非编程状态下不断扩充和完善知识库, 进而不断改善系统性能的能力, 即自学习能力。

(4) 智能软件系统的解释机制运用知识库中求解过程产生的各种中间结果, 能回答用户的提问, 给出求解过程的推理路径显示。这种解释机制提供了系统的一种透明界面, 加强了用户对系统的可接受性。

最后指出, 既然智能软件系统也是一个计算机程序, 从理论上讲, 可用任何语言编制程序: C 语言、PASCAL 语言、BASIC 语言。但是人们之所以常用 LISP、PROLOG、OPS83 等语言, 主要在于这些语言更适宜智能软件编程、处理符号、规则等, 也就是说, 所用语言不同, 构造智能系统的难易程度也大不相同。

1.2.3 智能技术实施的三个组成部分

智能技术实施于软件系统中有三个组成部分: 知识表示、知识推理、知识获取, 现分别介绍。

1. 知识表示

在通常程序中, 知识库级知识和控制级知识合二为一, 都体现在程序设计过程中。智能技术对此是分开的, 因而需要用适当的方法表示知识, 以便存储、检索、运用、增删和修改。所谓知识表示, 就是描述和组织知识的规则符号、形式语言和网络图等。其方法很多, 下面仅介绍产生式规则表示。

产生式规则可表示为

$$R_i: \text{IF } RLS \text{ THEN } RRS \quad (i=1, 2, \dots, N)$$

其中: R_i 称为规则集中第 i 条规则; RLS 是第 i 条规则条件部分, 可以是任何子句的逻辑组合; RRS 是第 i 条规则的结论部分, 可以是一个结论或操作, 或者是多个结论或操作的组合。

请看下列一组规则表示:

R1: IF 物体为绿颜色 THEN 它为农产品

R2: IF 物体包装在小容器内 THEN 它为精美食物

R3: IF 物体为冷冻食物 OR 农产品 THEN 它为易腐烂食物

R4: IF 物体重 15 磅 AND 为廉价物体 AND 不为易腐烂食物 THEN 它为常用食物

R5: IF 物体为易腐烂食物 AND 重 15 磅 THEN 它为火鸡肉

R6: IF 物体重 15 磅 AND 为农产品 THEN 它为西瓜

2. 知识推理

知识推理就是应用知识求解问题。知识推理过程就是问题求解的过程。知识推理技术就是使问题从初始状态转移到目标状态的方法和途径。知识推理的求解策略和推理模式作为控制级知识