

王耀生 著

智能预测系统设计与应用

人工智能



智能预测系统设计及应用

王耀生 著

气象出版社

内 容 简 介

本书是作者根据预测和人工智能特点,结合多年工作实践而写成的,智能预测系统是一个适应预测发展要求的应用系统,它包括两个库(知识库和智能数据库)和五个分系统(数据采集与质量管理,信息压缩与识别,信息合成与理解,推断与对策以及机器学习)。

本书在数据与有效信息的自动采集和提高整个系统的智能化水平方面进行了较系统的探讨和阐述,提出了一系列新的设想和做法。

本书可供气象、水文、地震、农业、林业、医学、军事、海洋、交通等业务部门参考,对科研、教学人员也将是一本适用的参考书。

智能预测系统设计及应用

王耀生 著

责任编辑 黄丽荣 李如彬

*

气象出版社出版

(北京西郊白石桥路46号)

北京昌平环球科技印刷厂印刷

气象出版社发行 全国各地新华书店经售

*

开本: 787×1092 1/32 印张: 4.75 字数: 105千字

1990年2月第一版 1990年2月第一次印刷

印数: 1-2500 : 定价: 3.50元

ISBN 7-5029-0330-5/TP·0018

前 言

自然科学在过去的一段漫长岁月里，呈现出学科分支愈来愈细的现象，从而把人们的视野引向更深入、更专门的领域。近代科学的发展出现了另一个新的趋向——多学科的综合，它的一个显著特点是在一些学科的边缘交叉点上，形成了新的“生长点”，遗传工程和知识工程的出现充分说明了这一点。

近些年来，人工智能作为一门边缘性学科，由于它的实用价值而日益引起人们的重视和广泛兴趣，专家系统是人工智能领域中发展比较迅速的一个分支，已经在实际工作中发挥了相当的作用。国际上比较有名的专家系统如DENDRAL用于帮助化学家分析化合物结构，RI是DEC.公司为用户配置计算机而专门设计的，MYCIN系统用于医疗诊断，PROSPECTOR则专门用于地质探矿。在我国，早期的中医专家系统是用于模拟一些名中医的诊断过程，取得了一定效果。进入80年代以来，气象、地震、地质、农业以及军事等部门相继开展了人工智能系统应用的研究，特别在气象部门，遍及全国的气象台站，目前在用专家系统预报灾害性天气方面取得了令人瞩目的成就。

虽然，在我国各种专家系统发展迅速，特别是预测领域，如天气预报、地震预报、产量预测和灾情预测等，但总的来说，现有的工作大都处于低水平阶段。这主要表现在以下几个方面：

① 基本数据、信息的输入是采用人读手键方式的，这不但增加劳动强度，降低工作效率，而且产生了相应的人为误差。

② 绝大多数系统的知识只是来源于专家，而且往往只是模拟了专家的一些静态的、结论性的知识，对动态的、关联性的知识还缺乏表述，尤其缺乏系统的自我修复和自我完善的能力。

③ 系统功能还不强，使用还不够灵活、方便。

数据采集和知识获取是两个既有区别，又有联系，既相互促进，又彼此制约的问题。如何实现数据采集的自动化和提高预测系统的智能化水平，以适应预测工作发展的需要，则是本书提出的系统设计所要解决的基本问题。

全书共分九章，第一章谈预测与人工智能，从分析这两个问题的特点入手，讨论了智能预报这一分支形成发展的必然性。第二章讨论知识库，探讨了预报知识的特点、来源和类型，着重讨论了知识的分层结构和表述方式。第三章对智能数据库的内容与构成，以及它与知识库的关系进行了阐述，指出我国资源库管理和布局的特点。第四章介绍数据接收和处理，对遥感数据与直接探测数据作了一些对比分析，指出它们各自具有的优势和不足，简要介绍数据处理的内容和方法。第五章对数据压缩和识别问题进行讨论，说明压缩与识别的必要性，介绍了具体的图形、图象识别的基本思路和方法。第六章讨论信息合成与理解，首先说明合成与理解的含义和所要达到的目的，然后举一个具体事例来加以说明。第七章是推断与对策，这里包含两部分内容，其一是预测推断，根据已有知识对多种信息综合理解，可能得出并非唯一的结论，然后根据需求差异、变化以及结论条件的可能

性作出相应的对策。另一是程序推理的搜索控制策略。第八章讨论的是机器学习，本书首先介绍了机器学习的两个基本概念：机器知识的来源和机器学习的含义，随后讨论了开发机器学习的阶段性和机器学习的两种基本方法，重点阐述了智能预测系统中机器学习的内容和遵循的两个基本原则，最后谈及当前机器学习中可能遇到的困难等。第九章介绍一个实用的气象预报专家系统，这个系统已经在18个省、市、自治区气象台或研究所推广应用，取得了一定的经济效益和社会效益。

本书的前八章组成了一个完整的体系——智能预测系统。为了使论述能比较具体，在各章的讨论中，较多的引用了气象方面的有关材料作为具体事例介绍，这丝毫也不影响它的通用性。

智能预测是一个既困难又很有吸引力的领域，它的前景十分迷人。作者希望这本书的出版能起到抛砖引玉的作用，在各位有志之士的共同努力下，将预测工作提高到一个新的水平。

目 录

前 言

第一章	预测与人工智能	(1)
1.1	预测问题	(1)
1.2	三种预测方法简介	(2)
1.3	人工智能问题	(5)
1.4	智能预测系统产生的历史必然性	(8)
1.5	智能预测系统的基本构成	(9)
1.6	智能预测系统的发展前景	(11)
第二章	知识库	(14)
2.1	气象预报知识的特点和现况	(14)
2.2	预测知识的类型	(15)
2.3	预测知识库的结构	(16)
2.4	知识表示问题	(18)
2.5	相似性概念在智能预测中的应用	(24)
2.6	小结	(26)
第三章	智能数据库	(28)
3.1	数据类型和数据结构	(28)
3.2	智能数据的基本特点	(30)
3.3	智能数据库结构	(31)
3.4	资源库的管理和布局	(36)
3.5	讨论	(40)
第四章	数据采集与质量管理	(41)
4.1	气象数据的来源和种类	(41)

4.2	气象数据的预处理	(43)
4.3	质量控制问题	(45)
4.4	直接探测与遥感数据的比较	(48)
4.5	气象数据处理出现的新情况	(50)
第五章	信息压缩与识别	(52)
5.1	问题的提出——气象探测数据和预报要求间的 差异	(52)
5.2	信息压缩的必要性和可能性	(53)
5.3	信息压缩	(55)
5.4	气象图象(形)的识别	(56)
5.5	小结	(68)
第六章	信息合成与理解	(69)
6.1	信息合成的含义	(69)
6.2	信息合成的困难	(70)
6.3	过去的工作及存在的问题	(71)
6.4	智能型信息合成的特点	(72)
6.5	信息合成的步骤和内容	(73)
6.6	理解	(75)
6.7	非数值模拟——一种新的试验方法	(77)
6.8	小结	(82)
第七章	推断与对策	(84)
7.1	形式推理方法	(85)
7.2	决断——预报集成问题	(91)
7.3	对策	(96)
7.4	小结	(99)
第八章	机器学习——知识自动获取	(101)
8.1	机器学习的基本概念	(101)
8.2	开发机器学习的阶段性	(103)
8.3	机器学习的两种基本方法	(104)

8.4	机器学习的内容和遵循的两个基本原则	(105)
8.5	机器学习所遇到的难题和可能的解决方案	(111)
第九章	介绍一个实用的气象预报专家系统	(114)
9.1	概述	(114)
9.2	气象科学的基本特点	(115)
9.3	系统的设计思想和原则	(119)
9.4	系统结构与功能	(121)
9.5	预报知识和知识表示	(124)
9.6	实时预报	(129)
9.7	知识获取——类比和归纳	(133)
结束语		(138)
参考文献		(140)
英文提要		(142)

第一章 预测^{*)}与人工智能

1.1 预测问题

所谓预测问题，实质上就是以过去的已知状况作为输入，在预测算子作用下，得到未来结果输出的过程（见图1.1）。

预测问题之所以一直是人们企图解决，而由于困难又没

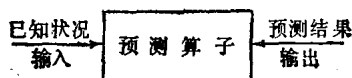


图1.1 预测问题示意图

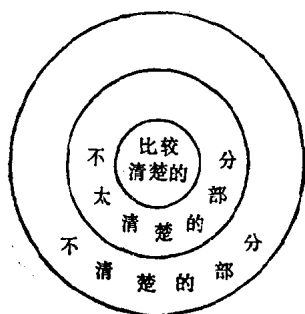


图1.2 预测知识现状图

^{*)} 预测和预报是同一个意思，本书中根据习惯，在不同地方采用不同的叫法。

解决的问题，是因为它存在着未知性。由图1.1可以看到，这种未知包括两方面：对过去的状况存在未知。对预测算子的未知。

过去的状况尽管是已经发生的事情，但已发生的不等于是人们认识的。已经记录的数据可能漏失了相当的信息，也可能提供某些错误的信息。预测算子的选择更是关系到对预测问题的认知，这与人们已有的知识有关。图1.2给出了预测知识的总体情况，它表明，在所有预测知识当中，只有一小部分人们比较清楚。

正是上述这些原因，使预测容易带有主观性，有人甚至把预测称之为“科学加艺术”。一般而言，预测不存在唯一和确定的方法，所谓的唯一确定，往往只是主观意义下的结果。应该说最佳预测实质上是最佳综合的同义语。

1.2 三种预测方法简介

1.2.1 动力预测方法

这是建立在牛顿力学基础上的一组相应的预测方程。在天文学中，日月食预报是从天体力学的质点三体运动方程出发的；地震动力预报是以地质力学方程组为基础的；在大气和海洋学的预报中，则以流体力学和热力学方程组为基本方程组。它们的共同基本形式是：

$$\frac{dA}{dt} = f(M, t) \quad (1.1)$$

式中 $\frac{d}{dt}$ 表示质点随时间的个别变化，A代表质点的某一物理属性的特征量，f是影响质点的外界条件，M代表空间坐标，t代表时间。

(1.1)式中如果 $f=0$ ，表示没有外界影响，则 $\frac{dA}{dt}=0$ ，即 $A=\text{const}$ ，表示该物理属性的特征量守恒。如果(1.1)式右边的 f 项是确定的，那么，它就成为一个严格的数学物理方程，在一定的初值条件下，它的解就存在并唯一确定了。根据数理方程要求，初值应该是方程在 $t=0$ 时刻的解，因此预测问题就转化为：如何使初值与方程相容，并使之满足 $t=0$ 解的条件。这就是1.1节中所说的“已经记录的数据”的影响，这在气象预报中叫做“初值化”，如何得到所需的预测结果。这一般采用时间逐步外推的数值求解方式。当然，对于三体运动这一类问题，可以采取更简捷的方法。

这类预测方法的基本前提是：质点的运动变化完全由(1.1)式的右端项决定，只要右端项确定，解就随之确定，这里不存在任何的不确定问题。可是当人们作进一步探讨时就会发现，恰恰是在(1.1)式右端项的确定过程中，除了一些简单的预测问题外，相当多的预测问题带有经验性和主观性。

1.2.2 统计预测方法

这类预测方法在社会科学和经济领域有着更广泛的应用，它是以概率统计为基础，统计量（平均值、方差等）为对象进行的预测，可以反映宏观的规律性，如果假设相关联系可以表征某种因果关系，则多元回归可看作多因素的一种联系。

这种方法对于一些情况复杂，满足统计特征条件，又需要进行趋势性预测的问题（如人口问题，经济与市场动向问题），采用这类方法是比较有效的。有些由于实际需要，而又没有其他更好的方法可用（如长期天气预报），应用这种方法也

是可行的。但对于非平稳过程，对一些小概率事件，如各种灾害的预测（暴雨、冻雹预报，地震预报，海啸预报等），统计预测方法显然会遇到一定的困难，它的效果必然会有局限。

近年来，不少人提出了动力与统计结合，走动力统计或统计动力的道路。前者可以用气象上的MOS(Model -Output Statistics)为代表，它利用动力预报的结果输出，再结合局地条件进行统计处理，作出具体的天气预报。后者则发展形成了一个独立的预测分支——随机动力预测方法。在数学上，也突破了原来的数学物理方程的框架，进入了一个新的、随机微分方程的领域。这方面的工作尽管已有一、二十年的历史，但由于遇到了物理上和数学上的困难，所以目前只在线性领域取得了一些进展。

1.2.3 经验预测方法

在自然界各种因素间具有错综复杂的联系，对其中的一些关系很难应用数量来表示，只能采用一些定性的描述，例如，一场重大灾害的发生和演变的条件、因素等，决不是用公式和方程式所能概括的。

同样地，对通常事件的预测，一般也不是动力或统计预测所能解决的。长期的实践充分证明，一个有经验的专家在处理这类问题时，要比一个初出茅庐的人高明得多。这些经验有编书出版的，也有在民间流传的。

这类预测方法的一个显著特点是没有固定的程式，它会因人、因事、因时、因地而异。

经验预测区别于前两种预测方法的另一个特点是，它以综合推理为主，非数值计算为其基本特征，这里需要作的是对已获取的信息进行综合分析，最后作出推断。

经验预测方法是以丰富的信息为前提，以经验性知识（技巧）为基础的。“心理学家已经指出，当象棋大师盯着一个棋位时，在他们的脑子里出现了几千盘重要的棋局”。由此可见，信息与知识情况是经验预测能否作好的两个必要条件。

1.3 人工智能问题

1.3.1 人工智能的一些基本概念约定

自1956年提出人工智能（AI）到现在已过去了三十多年，这期间经历了由理论探索到应用研究的巨大变化。正由于此，专家系统做为人工智能的一个前沿领域，已取得了相当的成就。

尽管有关人工智能的研究已有几十年的历史，但人们对它的含义还存在着不同的认识，目前尚无普遍公认的定义。最直观的理解是：“由计算机来表示和执行人类的智能活动就叫人工智能。”而比较有代表性的两种认识是：一种认为“人工智能是人类智能的模拟”，另一种则认为“人工智能是人类智能的放大”。

在人工智能科学中，引入了一系列概念，对这些基本概念认识做出统一约定，对工作和交流都是很有好处的。本书选用了如下约定：

约定一：数据是信息的形式化表示。

约定二：信息是宏观事物发展内在规律的外部表现。

约定三：知识是系统化了的的信息。

约定四：智能是获取、理解和运用知识的能力。

通过上述讨论，可以清楚地看到，智能活动的核心是知识，没有知识也就无所谓智能活动，因此可以说，人工智能研究的关键是知识的研究，这就构成了人工智能知识工程研

研究的三大基本问题。

① 知识表示：它要研究的是在计算机中要采用怎样的符号数据结构，才能恰当地表示相应的知识，以便于应用。

② 知识利用：如何利用已有的知识解决面临的问题，以及如何才能使推理迅速、有效，则是本课题所要解决的基本内容。

③ 知识获取：这是人工智能研究中最困难也是最重要的部分。它的基本内容是如何让计算机自动地获取为解决为题所必须的知识。它实际上要解决的是机器学习问题。

1.3.2 人工智能的研究和应用领域

人工智能学还是一门比较年轻的科学，要严格地明确它的研究对象和应用领域是有一定困难的，它又是一门边缘性学科，与其他学科在不少方面有着交叉。然而，人工智能毕竟是一门独立的学科，有着自己特定的研究课题：如自然语言处理、自动定理证明、智能检索、计算机视觉、智能程序语言、自动程序设计和问题求解等。需要说明的是这里所说的问题求解是狭义的，因为从广义来说，上述这些课题都可以叫做问题求解，这里的问题求解是指基于知识的、从有限的可能解空间中寻找一个或一组解的方法。

• 自然语言处理的一个重要应用领域就是机器自动翻译。目前已初步实现了几种西文之间文字的机器翻译，口语翻译由于语音识别的限制，只有局部的成功。至于中西文之间的翻译，由于文字和语法结构上的差异，尚需作更多的努力。

• 智能检索在数据库设计和管理使用中，有着巨大的应用价值，因为一般的数据库检索只是提供实况数据，不能也不可能得到推理性结果，而智能检索则可以利用知识，经过演绎，得出一般数据检索得不到的答案。

• 计算机视觉在机器人设计和遥感应用(特别在军事上)上都有十分重要的作用。它的主要工作特点就是从极其庞大、难以处理的输入数据中提取(识别)某些有意义的、易于处理的描述,这种描述往往是与特定的要求和知识密切相关,最终导致概念模型的建立。

• 问题求解是一个研究和应用都很广泛的课题,本书所讨论的核心——预测问题,就是一个问题求解,智能预测系统可看作是一个问题求解系统,它包括难题分解,若干子问题分层规划,解空间的搜索和回溯技术以及解的最佳组合等。

• 自动定理证明、智能程序语言和自动程序设计等,它们的题目已明确了相应的使用限定范围,故不再叙述。

需要指出的是上述划分只具有相对的意义,例如本书所述的智能预测系统,它还要涉及到计算机视觉(图象、图形识别)、智能检索、专家系统以及最佳对策等一系列人工智能问题,而且严格地说,这个系统就是在气象预报专家系统的基础上发展形成的。

1.3.3 智能科学与传统数学在问题求解处理上的差别

“传统的数学方法,是把客观事物场加以抽象,舍弃事物的内涵或语义,提取事物的数量变化规律,建立数学空间进行问题求解,得出具有唯一性的定理和结论。这种数学空间方法可以表征如下”(见图1.3)。

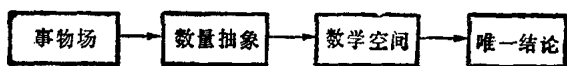


图1.3 传统数学的问题求解方法

智能学方法“不但分析事物的数量方面，更为主要的是分析事物的内涵或语义，建立知识空间，得出的命题并非唯一，亦即可能有多种答案或结论。这种知识空间方法可以表征如下*）（如图1.4）：

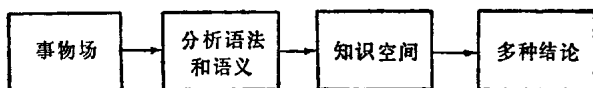


图1.4 智能学的问题求解方法

从图1.3和图1.4的对比中，可以看到，上述两种方法的差异表现在除第一框之外的全过程。首先是对客观事物的特征提取，经典数学方法考虑的是数量关系，智能法则考虑语义和内涵；随之数学方法是建立数学模型和方程，智能法则形成知识模型和规则；数学法强调的是解的存在唯一，而智能法恰恰突破了解的唯一性限制。正是由于智能法的这些特点，扩展了问题求解的领域。

1.4 智能预测系统产生的历史必然性

1.4.1 学科发展的需要

从预测的观点来看，上面介绍的三种预测方法中，各有自己的优势和不足，如何保持优点，克服不足，自然成为需要考虑的问题。特别是经验预测方法，大量的手工操作与信息的高速传输很不适应，经验性知识的客观表示和系统化则是另一个问题，这两个问题在各类预报专家系统的实践中已日益暴露出矛盾的尖锐性。在人工智能方面，过去知识工程的三大课题是分别加以研究的，尽管也都取得了相当的进展，获得了一些有实用价值的成果，但“难于在实际中应用”

*）引号中的两段引自王遇科“人工智能对气象预报技术的挑战”。