

智能情报系统

张玉峰 编著

武汉大学
出版社

智能情报系统

张玉峰 编著

武汉大学出版社

1991年·武汉

内 容 简 介

本书从理论、方法、技术和实践方面详细论述了智能情报系统这一新领域。全书共分十二章,内容包括:人工智能和智能情报系统的基本概念,知识表示;知识推理方法,智能情报系统的结构与建造,情报知识库,智能检索与推理机制,人机接口,知识获取和机器学习,以及智能情报系统实例等。

本书可作为大专院校科技情报及其它有关专业的本科生和研究生的教材或教学参考书,亦可供广大情报、图书、档案及人工智能应用领域的科技人员借鉴。

智能情报系统

张玉峰 编著

*

武汉大学出版社出版发行

(430072 武昌 珞珈山)

武汉大学珞珈高技术发展中心电脑排版

武汉大学印刷厂印刷

*

850×1168 毫米 1/32 11.25 印张 283 千字

1991 年 5 月第 1 版 1991 年 5 月第 1 次印刷

印数:1-1000

ISBN 7-307-00630-8/TP·35

定价:4.80 元

前 言

计算机情报系统产生于 50 年代,从脱机批处理,发展到交互式联机系统和网络系统,应用越来越广泛。

随着人工智能与专家系统这门新学科的发展,以及广泛的社会情报需求与供给的要求,智能情报系统应运而生。它将引起情报理论、方法和技术的一次革命,成为情报科学的新兴的重要研究领域。《智能情报系统》是系统介绍智能情报系统的基本理论、方法、技术和应用的一本专著。

本书系统总结了国内外及作者本人多年来关于智能情报系统的研究成果及实践经验,运用情报系统技术与人工智能技术相结合、理论与实践相结合的方法,全面论述了智能情报系统设计的新思想、新方法和新技术。全书分为三部分。第一部分由 1—3 章组成,主要介绍人工智能与专家系统的基本理论、方法和技术;第二部分由 4—8 章组成,详细论述智能情报系统的结构原理与建造技术,及其主要组成部分的结构、设计与实现方法;第三部分由 9—12 章组成,主要介绍几种典型的智能情报系统实例。本书第一章至第十章由张玉峰执笔,第十一、十二章由何皓执笔。

在编著本书过程中,得到了武汉大学自然科学研究处、武汉大学珞珈高技术发展中心和国内外许多朋友的热心帮助,在此一并深致谢意。

由于智能情报系统涉及的领域广,且处于发展阶段,加上我们的知识有限,编写经验不足,疏漏或错误之处在所难免,敬请专家和读者批评指正。

编著者:张玉峰

一九九一年三月于武昌珞珈山

目 录

第一章 人工智能与情报科学	1
§ 1.1 数据、情报与知识	1
1.1.1 数据	1
1.1.2 情报	1
1.1.3 知识的定义和类型	2
§ 1.2 计算机情报系统	4
1.2.1 计算机情报系统的结构	4
1.2.2 计算机情报系统的类型	5
§ 1.3 人工智能	7
1.3.1 基本概念	8
1.3.2 人工智能的基本内容	8
1.3.3 人工智能的重要应用	9
1.3.4 人工智能语言和工具	11
§ 1.4 专家系统	12
1.4.1 专家系统的特征	12
1.4.2 专家系统的结构	13
1.4.3 专家系统的作用和问题	15
§ 1.5 人工智能在情报科学中的应用	15
1.5.1 情报科学与人工智能的关系	15
1.5.2 人工智能在情报科学中的应用领域	16
§ 1.6 智能情报系统概述	18
1.6.1 智能情报系统的结构	19
1.6.2 智能情报系统与专家系统的区别	20
1.6.3 智能情报系统的发展	21
第二章 知识表示	23
§ 2.1 概述	23

§ 2.2 产生式规则表示法	25
§ 2.3 谓词逻辑表示法	27
2.3.1 谓词逻辑	27
2.3.2 情报知识的表示与检索	30
2.3.3 谓词逻辑表示的特点和问题	32
§ 2.4 语义网络表示法	33
2.4.1 图形表示	33
2.4.2 性质继承及语义网络的推理	35
2.4.3 语义网络的特点和问题	36
§ 2.5 框架表示法	37
2.5.1 框架的基本结构	37
2.5.2 附加过程	39
2.5.3 框架系统中的知识组织	41
2.5.4 框架系统的推理机制	42
2.5.5 文献的框架表示	43
2.5.6 框架表示法的评价	46
第三章 知识推理方法	48
§ 3.1 概述	48
§ 3.2 正向推理	50
3.2.1 正向推理的分析	50
3.2.2 解决冲突的策略	53
3.2.3 正向推理算法	54
§ 3.3 逆向推理与算法	54
§ 3.4 双向推理与算法	57
§ 3.5 不精确推理	58
3.5.1 不确定性	58
3.5.2 模糊逻辑	59
3.5.3 确定性理论	61
3.5.4 不精确推理与算法	63
第四章 智能情报系统的结构与建造	67
§ 4.1 智能情报系统的目标	67
§ 4.2 系统模型与功能	68

§ 4.3 智能情报系统的实现结构	71
4.3.1 分布式系统	71
4.3.2 元控制系统	76
§ 4.4 多知识源的处理	77
§ 4.5 智能情报系统设计方法	79
4.5.1 原型系统方法	79
4.5.2 知识获取过程	80
4.5.3 原型系统的设计与实现	82
4.5.4 情报系统设计准则	88
4.5.5 系统评价	90
§ 4.6 小型辅助检索专家系统的建造	91
4.6.1 问题分析	91
4.6.2 知识库的设计	92
4.6.3 系统实现过程	94
第五章 情报知识库	99
§ 5.1 智能情报系统中的知识类型	99
§ 5.2 情报知识库的概念和种类	100
§ 5.3 规则库	103
5.3.1 规则库的组成元素	103
5.3.2 规则库的结构	104
§ 5.4 专业知识库	105
5.4.1 叙词表的结构和功用	106
5.4.2 叙词表的构造方法	108
5.4.3 叙词表的知识表示	112
5.4.4 叙词表的实现方法	114
5.4.5 叙词表的管理	116
§ 5.5 文献知识库的结构与设计	117
5.5.1 文献数据库中的问题与解决途径	117
5.5.2 扩展的实体关系模式	118
5.5.3 文献知识库的设计	119
5.5.4 文献知识库的维护	125
§ 5.6 演绎数据库	125

§ 5.7 知识库的管理与发展	127
第六章 智能检索与推理机制	129
§ 6.1 智能检索概述	129
§ 6.2 智能检索模型	130
§ 6.3 检索推理机	132
6.3.1 规则推理机	132
6.3.2 框架推理机	134
6.3.3 近似推理机	135
6.3.4 逻辑演绎推理机	137
§ 6.4 智能检索处理	138
§ 6.5 不精确推理检索方法	143
6.5.1 检索模型	143
6.5.2 Dempster—Shafer 理论	146
6.5.3 检索算法	148
6.5.4 检索实例	149
6.5.5 结论与问题	152
§ 6.6 大型搜索空间的处理方法	153
6.6.1 搜索空间的分解	153
6.6.2 搜索空间的抽象	154
§ 6.7 联想检索	156
6.7.1 联想检索的基本原理	156
6.7.2 启发式知识	157
6.7.3 检索算法	158
6.7.4 检索实例	160
§ 6.8 分类检索	161
6.8.1 分类检索原理	162
6.8.2 分类检索步骤	163
6.8.3 检索实例	164
§ 6.9 继承检索	166
6.9.1 继承检索原理	166
6.9.2 继承检索算法	167
6.9.3 多重继承问题	169

第七章 智能情报系统的人机接口	171
§ 7.1 认知理论在人机接口中的作用	171
§ 7.2 智能接口	173
7.2.1 智能接口的定义和特征	173
7.2.2 友好用户的标准	174
7.2.3 智能接口的主要任务	174
7.2.4 智能接口的结构与设计	178
§ 7.3 用户模型	181
7.3.1 概述	181
7.3.2 用户模型的分类	182
7.3.3 建立用户模型的主要任务	184
7.3.4 构造用户模型的方法	186
7.3.5 用户模型在情报系统中的应用	192
§ 7.4 自然语言处理的主要方法	193
7.4.1 语法分析	193
7.4.2 语义结构	196
7.4.3 语用学	197
第八章 知识获取和机器学习	199
§ 8.1 知识获取	199
8.1.1 知识获取的主要方法	199
8.1.2 知识获取中的问题及解决办法	202
§ 8.2 机器学习	204
8.2.1 概述	204
8.2.2 学习系统的模型	205
8.2.3 机器归纳学习	207
§ 8.3 机器归纳学习系统: AQ15	212
8.3.1 系统概述	212
8.3.2 系统结构和学习算法	213
8.3.3 样例和系统评价	217
§ 8.4 机器学习在情报系统中的应用	219
8.4.1 情报系统中的知识获取任务	219
8.4.2 具有学习功能的情报系统	221

8.4.3	概念学习	222
8.4.4	文献描述的遗传学习算法	225
第九章	智能中间人系统:CANSEARCH	230
§ 9.1	情报检索中间人	230
§ 9.2	专家系统模型分析	231
§ 9.3	CANSEARCH 系统设计	232
9.3.1	CANSEARCH 系统的领域和知识	232
9.3.2	抽象查找空间	234
9.3.3	菜单层次结构	234
§ 9.4	规则系统的实现	237
9.4.1	规则、黑板和语境	237
9.4.2	用 PROLOG 实现系统	240
§ 9.5	CANSEARCH 系统的检测	241
9.5.1	检测目标	241
9.5.2	检测提问	242
9.5.3	检索表达式的产生	242
9.5.4	检索表达式的质量评估	243
9.5.5	检索结果的相关性评价	245
§ 9.6	结论	247
第十章	智能情报检索系统:CODER	248
§ 10.1	CODER 计划背景	248
10.1.1	情报检索中的问题与研究	248
10.1.2	CODER 系统概述	249
§ 10.2	CODER 系统结构	251
§ 10.3	知识库	254
10.3.1	综合型文献知识库	254
10.3.2	人工智能专业知识库	256
10.3.3	字典库	257
§ 10.4	文本分析	258
10.4.1	文本分析方法	258
10.4.2	文献分析实例	260
§ 10.5	扩展的布尔检索	263

10.5.1 检索策略的研究	263
10.5.2 扩展检索模型	265
10.5.3 p-norm 检索专家	267
10.5.4 检索结果	269
§ 10.6 用户接口	271
§ 10.7 结论与问题	273
第十一章 咨询专家系统: PLEXUS	274
§ 11.1 PLEXUS 计划	274
11.1.1 目的和目标	274
11.1.2 计划背景	276
§ 11.2 PLEXUS 的设计	278
11.2.1 硬件/软件	278
11.2.2 系统研制的步骤	278
11.2.3 原型评估	280
§ 11.3 PLEXUS 的组成	281
11.3.1 知识库	281
11.3.2 用户接口	285
11.3.3 推理机	287
11.3.4 体系结构	287
§ 11.4 PLEXUS 系统的工作机理分析	288
11.4.1 PLEXUS 系统的主要问题	288
11.4.2 问题表达式的构造	289
11.4.3 处理未知的输入词	292
11.4.4 构造布尔检索表达式	293
11.4.5 修改检索表达式	294
§ 11.5 PLEXUS 咨询处理过程	296
§ 11.6 PLEXUS 遇到的难题	301
11.6.1 拼写错误与词义模糊	301
11.6.2 分类法的使用	303
11.6.3 主题领域的扩充	304
§ 11.7 一个检索实例	305
第十二章 编目专家系统	312

§ 12.1 编目基础	312
12.1.1 编目及其方法	312
12.1.2 计算机编目与编目规则	314
§ 12.2 编目专家系统的实用性研究	316
12.2.1 戴维斯(Davies)和詹姆士(James)的实验	316
12.2.2 其他的编目专家系统研究	322
12.2.3 编目专家系统的效率与实用性	325
§ 12.3 编目专家系统的建立	331
12.3.1 问题的提出及建立方法	331
12.3.2 编目专门知识及来源	333
12.3.3 启发式证据的研究	334
12.3.4 编目及知识表达方法	340
§ 12.4 我国编目工作的现状及发展	341
智能情报系统附录	343
主要参考文献	346

第一章 人工智能与情报科学

§ 1.1 数据、情报与知识

数据、情报和知识是情报系统处理的基本对象,它们描述了客观世界,互相关联、互相交错,在使用过程中互相转换。

1.1.1 数据

数据是表示情报和知识的数字、字符或符号的集合。数据是对世界中客观事物直接观察的结果,它被看作自然对象或原始材料。一个数据元素是一个实体和与其相关的值,如统计数字、实验结果、文献原文、图象等。数据元素的集合就组成了数据库。

1.1.2 情报

关于情报的含义,专家、学者从不同角度给出了许多种定义。目前国内外对情报概念的看法还不一致,下面列举一些有代表性的观点。

苏联情报学家 A. И. 米哈依洛夫认为:“情报是作为存储、传递和(或)转换的对象的知识。”这是知识类情报定义的典型。

美国学者斯拉麦卡(V. Slamecka)认为:“情报就是有用的数据或被认为是有用的数据。”他强调情报的有用性。

许多人认为计算机情报系统的输出等于情报,但质量不高。我们可抓住情报现象的本质认识情报概念。情报具有精确性、及时性和相关性。精确性,意即情报正确、清楚,且准确地反映了数据的意

义,避免偏见,又使用户易于理解。及时性,就是用户能在他们的需求时间内得到情报。相关性,是情报的关键特性,即情报仅相关于某接受者或某种用途而存在。

情报是由系统(或人)感知和(或)产生的知识。数据被感知和利用后就变成情报。情报必须变成系统或人求解问题的知识结构的一部分,否则,实际问题就不能被解决。

情报的基本目的是产生和扩展知识。情报具有以下应用特性:情报可作为理解过去和目前状况、预言未来发展的基础;情报是形成知识的源泉和材料;在领域活动中,情报被用于扩展知识;情报的评估可创造新的知识;情报可被模型化,提供获取知识的结构。

1.1.3 知识的定义和类型

什么是知识,目前尚无统一的观点。这里将从人工智能研究者的角度列出几种有代表性的说法。

Bernstein 认为:知识是由特定领域中的描述、关系和过程组成的。描述是关于领域中对象及对象类的描述,它是用某种语言表达的。这种语言的基本元素是事物的性质或概念。关系是指对象之间的依赖性和关联性。典型关系有分类学关系(等级关系)、定义性关系和经验性联想关系(如医生看病所使用的因果模型)。过程是在推理或问题求解时规定要执行的操作。

人工智能专家 Feigenbaum 认为“知识是经过削减、塑造、解释、选择和转换的信息”。知识是一种人类设计的人工品。设计则是情报的再利用技术,即把一组离散的、不太直接相关的做法塑造成普遍接受和采用的系统原则。任何经过设计加工的知识,都会在大脑中“编辑”为一模型,如系鞋带,最初必记每一个步骤,经过一段反复执行,就形成“系鞋带知识”模型。同理,人类的知识经过工程化,将来都会经过编辑送入智慧型机器。

按知识的一般分类,任何领域知识都包括一般性知识(或公开知识)和专门知识(或称个人知识)两类。一般性知识包括公开

出版的定义、事实和理论,这些往往已收录在教科书和参考书中。专门知识主要是指专家或实践工作者多年积累的解决专业实际问题的经验知识和技能,或称启发式知识(Heuristic Knowledge),其含义是关于专家个人解决问题的决策行为的设想。它可以是未被出版的、经验的、非正式的、判断性的、且常常是过程式的,如实践经验的规则。这种知识能使专家作出有根据的猜测或推断,找到最有希望的解决问题的途径和有效地处理错误或不完全的数据。启发式知识主要来自专家或实践工作者,如分类、编目专家、情报检索专家、检索中间人以及用户。它还可来自书本,如高次文献、手册与内部资料。

从知识系统知识表示的角度来看,更侧重于考虑以下几类知识:

1. 有关对象及概念的知识

最基本的知识首先在于对事物的描述。对于事物来说,有客观存在的事物与想象中存在的事物之分。前者称为物理对象,后者称为概念对象。事物的描述,就是用已知事实的描述来有效地表示对象(或概念)的性质及它们之间的关系。

2. 有关事件、行为及状态的知识

这类知识包括与社会的事件相关的知识,与客观的或精神的行为相关的知识,与世界的状态相关的知识等。一般,这种知识是对于多个对象间的关系及其相互作用的描述。

3. 关于过程的知识

为了描述实现某任务或目标的一系列行为,如汽车驾驶,需要有启动条件、顺序关系、因果关系等有关的知识。该类知识是任务性知识,直接用于解决问题。

4. 元知识(Meta-Knowledge)

元知识描述并指示如何使用知识,即关于如何运用已知知识的知识。例如,为解决一个特殊任务而需完成的活动计划、组成和选择方面的知识等。这些知识在人类活动中起指导作用。元知识

主要用于知识系统的控制、推理、解释、知识获取和学习等方面,协调控制整个问题的解决。

知识系统的知识库可能包含的各类知识如图 1.1 所示。

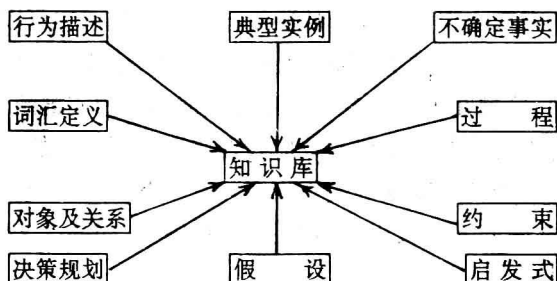


图 1.1 知识类型示意图

§ 1.2 计算机情报系统

1.2.1 计算机情报系统的结构

计算机情报系统,是利用计算机模拟手工情报处理过程,实现情报的存储、处理和利用的自动化系统。所有的计算机情报系统包含五个基本模块:输入、控制、处理、数据库和输出(图 1.2)。这些模块可取不同的结构、内容和工作方式,可是简单的或是非常复杂的。根据不同的应用目标可将这些基本模块组成所需要的情报系统。下面对这些模块作以简单解释。

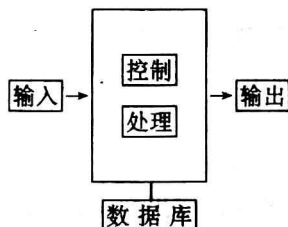


图 1.2 情报系统的结构

1. 输入模块

利用合适的方法和设备获取数据、原文、声音及图象,并输入情报系统。输入包括:转换、请求、询问、指示和发信号等。

2. 控制模块

它控制数据的输入和处理,并以不同方法产生需要的结果或输出。它控制和执行确定的操作,将数据转换成所需要的情报。控制机制常须考虑组织操作的特性、管理策略、用户的需求和系统的执行效能等。

3. 处理模块

它利用优化的方法和技术执行确定的操作:获得输入、存储与检索数据、处理与传递输出。处理部分要仔细考虑容量、时间、复杂度和计算等问题。

4. 数据库

数据库用来存储用户所需要的数据。数据库包括物理组织和逻辑组织两部分。物理数据库是物理存储器(如磁带、磁盘等)中的数据组织形式。逻辑部分确保及时的、相关的、准确的数据存取,它涉及系统的软件元素及数据的逻辑关系、存储结构及处理技术。

5. 输出模块

情报系统的输出是符合用户要求的情报或参考文献。输出精制加工技术包括:过滤、控制、询问、调整变化、模式化及决定策略等。

此外,情报系统还涉及人力的和设备的环境条件及其它设计要求,如综合性、竞争力、系统要求、用户接口、情报的质量和可用性、数据处理的要求、消费和效率要求及灵活度要求等。

1.2.2 计算机情报系统的类型

情报系统广泛应用于图书馆、各行业的情报机构、工商企业管理部门及政府机关等。一般来说,一个情报系统具有一个主要目标,如管理、决策、检索及分析等。而信息的存贮与检索是所有情报