

李锋刚 著

基于案例推理的 智能决策技术

Intelligent
Decision



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
安徽大学出版社

基于案例推理的智能决策技术

李锋刚 著



北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
安徽大学出版社

图书在版编目(CIP)数据

基于案例推理的智能决策技术/李锋刚著. —合肥:安徽大学出版社,2010.11
ISBN 978-7-81110-877-4

I. ①基... II. ①李... III. ①智能决策—决策支持系统 IV. ①TP399

中国版本图书馆 CIP 数据核字(2010)第 204740 号

基于案例推理的智能决策技术

李锋刚 著

出版发行:北京师范大学出版集团

安徽大学出版社

(安徽省合肥市肥西路3号 邮编 230039)

www.bnupg.com.cn

www.ahupress.com.cn

印刷:合肥创新印务有限公司

经销:全国新华书店

开本:169mm×228mm

印张:9

字数:157千字

版次:2011年1月第1版

印次:2011年1月第1次印刷

定价:23.00元

ISBN 978-7-81110-877-4

责任编辑:赵小文

装帧设计:李军

责任印制:陈如韩琳

版权所有 侵权必究

反盗版、侵权举报电话:0551—5106311

外埠邮购电话:0551—5107716

本书如有印装质量问题,请与印制管理部联系调换。

印制管理部电话:0551—5106311

前 言

案例推理技术是人工智能领域中的一种重要的基于知识的问题求解和学习方法。它解决问题是通过重用或修改以前解决相似问题的方案来实现的,是由目标案例的提示而得到历史记忆中的源案例,并由源案例来指导目标案例求解的一种策略,也是一种重要的机器学习方法。案例推理研究方法源自人类的认知心理活动,缓解了常规的知识系统中知识获取的瓶颈问题,它将定量分析与定性分析相结合,具有动态知识库和增量学习的特点。

1977年,美国耶鲁大学 Roger Schank 在研究认知科学的过程中,提出了“知识表示脚本”的概念。1982年,在《Dynamic Memory》中提出的以记忆组织包为核心的动态记忆理论,被认为是人工智能领域中最早的关于案例推理的思想。1995年召开了第一届国际案例推理大会,其后每两年召开一次。近年来国际上的研究性开发项目和应用系统不断增加,出现很多使用案例推理方法的商品化工具。人们对案例推理的研究和应用,已经深入到热门的电子商务、WWW、信息服务、规划、设计、诊断、分类、音乐、软件复用、医学、辩论、争端调解、智能教学以及辅助决策等领域。

智能决策技术是现代管理中的关键技术,涉及人工智能、管理科学、信息科学、知识系统工程、计算技术、运筹学和决策科学等,是由新兴学科的相互交叉、相互渗透而产生的新技术、新学科。它研究如何提高管理过程的智能水平及智能决策中的设计理论、方法与实现技术。近年来,决策技术与优化技术紧密结合、决策技术与计算智能紧密结合的趋向日益增强。

基于定性推理的智能决策技术、基于粗糙集的智能决策技术、基于证据理论的智能决策技术、基于群体层次分析的智能决策技术及基于 Multi-Agent 的智能决策技术的研究,极大地促进了智能决策技术的发展。这些技术虽然从不同角度、在不同程度上解决了决策支持系统中的一些问题,但仍然难以较好地解决以下问题:相关领域知识缺乏,专家通过经验解决问题,存在大量例外于规则的事实,难以建立模型,存在半结构化和非结构化问题等。而解决上述问

题正是案例推理技术的优势之所在,因此,本书研究了基于案例推理的智能决策技术 CBR-IDT。

本书共分 7 章,主要研究了基于案例推理的智能决策技术。第 1 章介绍了优化技术与智能决策及其发展的概况;第 2 章主要研究了案例型智能决策系统的构成及其框架;第 3 章分析了基于熵、遗传算法和主成分分析的属性选择的策略;第 4 章包括案例推理中的多策略相似性检索技术;第 5 章是基于禁忌搜索技术的案例推理技术;第 6 章给出一个基于案例推理技术的中医智能诊断系统。本书也是关于我国“三大显学”之一——“徽学”中新安医学在中医药现代化、智能化、信息化方面的探索和尝试。

本书的出版是在过程优化与智能决策教育部重点实验室条件支撑下完成的,是在安徽省自然科学基金项目“基于联系发现和启发式的优化型案例决策技术及应用研究(090416246)”、安徽省教育厅自然科学基金项目“新安医学防治中风病组方用药规律研究(2006KJ387B)”和安徽省高等学校青年教师科研项目“新安医学防治中风病的文献研究(2005jq1097zd)”资助下展开研究工作所取得的成果。

本书在撰写过程中,合肥工业大学管理学院倪志伟教授给予了悉心的指导;中国科学技术大学计算机学院陈恩红教授给予了极大的关心;安徽中医学院中医临床学院邵峦博士对相关内容给出了有益的建议;在此谨向他们表示诚挚的谢意。

本书在编写过程中,参考了大量国内外有关研究成果,笔者在参考文献中尽量列出,在此对涉及的专家和研究人员表示衷心的感谢,如有疏漏之处还请见谅。

由于作者水平有限,书中的错误或不妥之处在所难免,诚恳希望同行和读者批评指正,以便今后改正和完善。(E-mail:lifecan@sina.com)

作 者

2010 年 10 月

目 录

第 1 章 绪论	1
1.1 研究背景	1
1.1.1 决策与智能决策系统	1
1.1.2 优化技术与优化算法	2
1.1.3 智能优化算法与启发式方法	4
1.2 国内外研究现状	5
1.3 研究意义和内容	7
1.3.1 研究意义	7
1.3.2 主要研究内容和结构安排	7
第 2 章 基于案例推理的智能决策系统框架	10
2.1 案例推理技术	10
2.1.1 案例推理技术的起源和发展	10
2.1.2 案例推理技术的基本思想	11
2.1.3 案例推理技术的特点	13
2.2 基于案例推理的决策技术	15
2.2.1 决策案例的表示形式	15
2.2.2 决策案例的索引	17
2.2.3 决策案例的组织与检索	19
2.2.4 决策案例的修正	19
2.2.5 决策案例的学习和归纳	21
2.2.6 决策案例库的维护	22
2.2.7 案例推理中的 K 近邻技术	22
2.3 基于案例推理的智能决策系统框架	24

2.3.1	决策支持系统的发展	24
2.3.2	决策支持系统的基本结构	24
2.3.3	CDSS 系统结构	26
2.3.4	实现系统的关键技术	31
2.4	本章小结	34
第 3 章	CBR-IDT 中属性的优化选择	36
3.1	属性选择策略概述	36
3.1.1	研究回顾	36
3.1.2	属性选择的搜索策略和评价策略	37
3.1.3	属性选择的方法	37
3.1.4	属性选择的形式化	38
3.2	基于熵的属性优化选择	39
3.2.1	基于熵的属性约简模型和方法	39
3.2.2	性能评价策略	40
3.2.3	实验和结果	41
3.3	基于遗传算法的属性优化选择	44
3.3.1	遗传算法	44
3.3.2	基于遗传算法和相关性评价的属性选择	45
3.3.3	实验结果和分析	47
3.4	基于主成分分析的属性优化选择	48
3.4.1	总体主成分的定义	49
3.4.2	总体主成分的性质	50
3.4.3	标准化变量的主成分	51
3.4.4	样本主成分	52
3.4.5	实验结果和分析	53
3.5	本章小结	55
第 4 章	CBR-IDT 中多策略相似性检索技术	57
4.1	CBR-IDT 中案例的检索	57
4.1.1	案例检索的类型	57
4.1.2	案例检索的任务	57
4.1.3	案例检索的启发式方法	58

4.2 CBR-IDT 中案例相似性的度量	58
4.2.1 相似性度量概述	58
4.2.2 案例间相似性的度量	60
4.2.3 案例属性间相似性的度量	62
4.2.4 两种案例相似性度量方法	63
4.3 基于提升和投票的多策略相似性检索技术	65
4.3.1 为什么提升和投票	65
4.3.2 提升和投票	68
4.4 本章小结	71
第 5 章 基于禁忌搜索算法的案例推理技术	72
5.1 禁忌搜索算法	72
5.1.1 禁忌搜索算法概述	72
5.1.2 禁忌搜索的基本思想	73
5.1.3 禁忌搜索算法的基本流程	73
5.1.4 影响算法结果的要素	76
5.1.5 禁忌搜索算法的应用	76
5.2 禁忌搜索算法的关键参数	77
5.2.1 邻域及邻域搜索	77
5.2.2 禁忌表和禁忌表的大小	79
5.2.3 短期记忆和长期记忆	79
5.2.4 改变搜索路径的方法和期望指标	80
5.2.5 停止准则	81
5.2.6 搜索的效率	81
5.3 禁忌搜索算法求解 TSP 问题	82
5.3.1 TSP 问题的描述	82
5.3.2 朴素禁忌搜索算法求解 TSP 问题	83
5.3.3 基于启发式初始解的禁忌搜索算法求解 TSP 问题	85
5.4 基于禁忌搜索算法的图结构案例检索技术	88
5.4.1 相似性度量	88
5.4.2 禁忌搜索过程	89
5.4.3 一个两阶段的案例检索	91
5.5 禁忌搜索算法的改进策略	92

5.6 本章小结	94
第6章 基于 CBR-IDT 的智能决策系统	95
6.1 概述	95
6.1.1 中风病及新安医学	95
6.1.2 医学诊疗系统和中医诊疗系统的研究现状	96
6.1.3 研究目的和意义	97
6.2 案例推理智能决策系统	99
6.2.1 传统智能诊疗决策系统的缺陷	99
6.2.2 案例推理智能决策系统及其优点	99
6.2.3 基于案例推理的中风病诊断的一般过程	101
6.3 诊疗系统的设计	102
6.3.1 概念设计	102
6.3.2 结构设计	108
6.3.3 功能特点	110
6.4 系统实现	111
6.4.1 案例数据准备	111
6.4.2 系统实现过程	112
6.4.3 系统运行界面	116
6.5 本章小结	119
第7章 总结与展望	120
参考文献	124
附录	135
附录1	135
附录2	136

第1章 绪 论

1.1 研究背景

1.1.1 决策与智能决策系统

1. 决策

所谓“决策”，是指人们为了达到一定的目标，而决定行动方案并付诸实施的过程。在一定的人力、设备、材料、技术、资金和时间等因素的制约下，人们为了实现特定目标，而从多种可供选择的策略中做出决断，并付诸实施以求得最优或较好结果的过程就是决策。

决策的基本要素是决策者和决策对象，两者构成一个矛盾对立的统一体——决策系统(Decision System, DS)。决策者和决策对象相互作用的最一般的抽象是信息，信息是决策的必要条件，也是决策的基本要素。此外，决策活动还离不开决策者的决策技术与方法，以及最后获得的决策结果，包括行动方针、行动原则和行动方案等。

一个决策问题，须具备以下5个条件：

(1)决策目标。有一个希望达到的明确目标，即决策者对决策问题所希望实现的目标，可以是单个目标，也可以是多个目标。用决策准则或最优值 M 表示。

(2)自然状态。采取某种决策方案时，决策环境客观存在的各种自然状态。自然状态可以是确定的或不确定的，也可以是离散的或随机的。

一般来说，要求：①能给出系统所处各种可能的状态。用状态集 $S=\{S\}$ 表示，它是所有状态构成的集合，是随机变量。②能知道系统各种状态出现的可能性大小。用状态存在的概率 $P(S)$ 表示。

(3)决策准则。实现决策目标而选择行动方案所依据的价值标准和行为标准。一般来说，决策准则依赖于决策者的价值倾向和偏好，用 $C=\{C\}$ 表示。

(4)行动方案。实现决策目标所采取的具体措施和手段,要有多个备选方案。方案又称为决策变量,用 A 表示,所有方案构成的集合称为方案集,用 $A = \{A\}$ 表示。

(5)条件结果值。采取某种行动方案在不同自然状态下所出现的结果,能估算出系统在不同状态下的结果或效益。用受益值、损失值或效用值 $V(A, S)$ 表示,它是状态变量 S 和决策变量 A 的函数。

决策问题通常比较复杂,须采用抽象办法,找出参与决策过程诸变量之间的约束关系,建立数学模型。据此,一个决策问题的数学模型,可以表示为如下要素构成的 5 重数据结构:

$$D = \{M, \{S\}, P(S), \{C\}, \{A\}, V(A, S)\}$$

2. 智能决策系统

传统决策支持系统(Decision Support System, DSS)有时不能有效地处理现实生活中大量存在的半结构化或非结构化问题,这是因为它们很难用数学模型进行定量分析;有时虽然可以用模型来描述和求解,但很难取得基础数据;有些问题虽有多种模型可供选择,但对不同模型的参数条件和使用范围难以确定,特别是当实际问题与模型条件不一致时,无力指导用户修改模型。一般解决的办法是在传统决策支持系统中增加一个知识库,用来存放专家的决策思维过程、成功经验和有关科学知识等,从而构成智能决策支持系统(Intelligent Decision Support System, IDSS)。

1.1.2 优化技术与优化算法

智能优化技术通常都是基于智能优化算法的。20 世纪 80 年代以来,一些新颖的优化算法,如人工神经网络、遗传算法、进化规划、模拟退火、禁忌搜索、蚁群搜索、粒子群算法及其混合优化策略等,通过模拟或揭示某些自然现象或过程而得到发展,其思想和内容涉及数学、物理学、生物学、人工智能、神经科学和统计力学等方面,为解决复杂问题提供了新的思路 and 手段。这些算法独特的优点和机制,引起了国内外学者的广泛重视,并掀起了该领域的研究热潮,使其在诸多领域得到了成功应用。在优化领域,由于这些算法构造的直观性与自然机理,通常被称作智能优化算法(Intelligent Optimization Algorithms),或称元启发式算法(Meta-Heuristic Algorithms)。

所谓优化算法,其实就是一种搜索过程或规则,它是基于某种思想或机制,通过一定的途径或规则得到满足用户要求的问题解决方案。

按优化机制与行为而分,目前工程中常用的优化算法主要可分为:经典算

法、构造型算法、邻域搜索算法、基于系统动态演化的算法和混合型算法等^[1]。

(1)经典算法。包括线性规划、动态规划、整数规划和分支定界等运筹学中的传统算法,该类算法的计算复杂性一般很大,适于求解小规模问题,在工程中往往不是很实用。

(2)构造型算法。用构造的方法快速建立问题的解,通常算法的优化质量差,难以满足工程需要。譬如,调度问题中的典型构造型方法有:Johnson法、Palmer法、Gupta法、CDS法、Daunebring的快速接近法、NEH法等。

(3)邻域搜索算法。从任一解出发,通过对其邻域的不断搜索和对当前解的替换来实现优化。根据搜索行为,又可分为局部搜索算法和指导性搜索算法。

①局部搜索算法。利用局部优化策略在当前解的邻域中贪婪搜索,如只接受优于当前解的状态作为下一解的爬山法;接受当前解邻域中的最好解作为下一解的最陡下降法等。

②指导性搜索算法。利用一些指导规则来指导整个解空间中优良解的探索,如模拟退火(Simulated Annealing, SA)、遗传算法(Genetic Algorithms, GA)、进化规划(Evolutionary Programming, EP)、演化策略(Evolutionary Strategies, ES)和禁忌搜索(Tabu Search, TS)、蚁群优化(Ant Colony Optimization, ACO)、粒子群优化(Particle Swarm Optimization)等。

(4)基于系统动态演化的算法。将优化过程转化为系统动态的演化过程,基于系统动态的演化来实现优化,如神经网络和混沌搜索等。

(5)混合型算法。是指将上述各算法从结构或操作上相混合而产生的各类算法。这类算法的融合通常在机制上有一定的要求,即具有互补性。

当然还可以从别的角度对优化算法进行分类,如组合优化和函数优化,确定性算法和不确定性算法,局部优化算法和全局优化算法,离散型优化和连续型优化等。

在上述所述的智能优化中,有一大类算法具有拟物和仿生的特点,其思想大多来源于自然界,是对自然界中自主优化现象的模拟^[2]。如将金属降温、退火过程中的自然规律引入到优化求解中,产生了模拟退火算法;将生物种群进化的自然规律引入到优化求解中,产生了遗传算法;研究蚁群的个体间信息传递方式与作用规律时,产生了蚁群算法;对鱼群寻食和回游规律的研究,产生了鱼群算法;对鸟群扑食过程的观察和研究,促成了粒子群算法的产生;将复杂大系统中混沌规律引入到优化求解中,产生了混沌优化方法。此外,这类算法通常都还具有聚集性和弥散性的特点。

1.1.3 智能优化算法与启发式方法

有一大类有趣的问题,对它们的求解还没有快速、合理的算法,其中许多问题都是在应用中经常碰到的优化问题。对一个给定的比较困难的优化问题,我们通常很难找到有效的算法来得到其近似最优解,那么就可以采用一些随机算法来处理它。这些算法虽然不能保证获得最优解,但随机地选择充分多个“验证”后,错误概率可能会降低到我们满意的程度。

对许多实际的优化问题,已经有了一些高质量的算法。例如我们可以用模拟退火来解决 VLSI 设计中布线方案及组件放置问题,或者是货郎担问题。另外,许多 NP 难解的组合优化问题都能在现代的计算机上用 Monte Carlo 技术获得近似解。

一般来说,任何难解的任务都可被看成是要解决一个问题,反过来,它也可以被看成是对潜在解空间的一种搜索。因为我们最终的目的是要获得“最好”解,所以可以把这种任务看成是一个优化的过程。对小空间,经典的穷举法就足够了;而对大空间,则需要使用特殊的人工智能技术。

鉴于实际工程问题的复杂性、约束性、非线性、多极小值点、建模困难等特点,寻求一种适合于大规模且具有智能特征的算法策略,已成为有关学科的一个主要研究目标和引人注目的研究方向。

一个实际的智能系统,它的问题求解空间一般很大,空间中每一个结点对应于一个子目标,当求解每个子目标时,可供使用的知识有多个,仅靠盲目穷举的搜索方法很难奏效^[3]。为了提高求解效率,通常需要定义一种能反映求解路径优劣程度的数值因数 $f(\cdot)$,作为比较选择的参考,通常称该函数为评价函数(或估计函数, Evaluation Function)。

用评价函数来引导搜索的一般理论是由 P. E. Hart、N. J. Nilsson 和 B. Raphael 等人提出的。因为确定评价函数所依据的知识往往是经验的、直观的所谓启发式知识,因此这种搜索方法又被称为启发式搜索法(Heuristic Search)。

案例推理技术中使用启发式方法的主要表现在:①在案例求解的过程中使用领域专家的知识,如相似性案例的选择,案例的适配过程,求解结果是否成功等;②当问题空间比较大时,使用启发式函数可以帮助减小搜索范围,提高搜索效率。

案例主要由 3 部分构成:

(1)问题或情景描述:案例发生时要解决的问题及周围世界的状态。

问题或情景描述是对要求解的问题或要理解的情景的描述,一般包括:案

例推理发生时推理器的目标,完成该目标所涉及的任务,周围世界或环境与可能解决方案相关的所有特征。

(2)解决方案:问题的解决方案。

解决方案的内容是问题如何在某种特定情形下得到解决。可以是对问题的简单解答,也可以包括求解过程。

(3)结果:执行解决方案后导致的结果(周围世界的新变化)。

结果记录了实施解决方案后是成功还是失败的情况。案例推理技术给出建议解时,既可以是成功的案例,也可以是失败的案例(以避免产生不希望发生的结果)。当对问题缺乏足够了解时,在案例表示上加上结果部分能取得较好效果。

1.2 国内外研究现状

案例推理技术(Case-Based Reasoning, CBR)^[4]是与以下概念紧密相关的,如基于范例的推理,基于事例的推理,基于实例的推理,基于示例的学习,基于案例的设计等。这些概念之间或者在内涵和外延之间略有差异,或者是从不同的角度上来看待或应用此项技术。

现有关于案例推理技术的研究,基本可以分为两大类:一类是基本案例推理技术的研究,或称为朴素案例推理技术的研究;另一类是泛案例推理技术的研究,或称为扩展案例推理技术的研究。第一类包括以 K-NN 技术为核心的案例推理技术,第二类主要是案例推理技术与其他方法的融合,如贝叶斯方法,粗糙集方法,模糊集方法等。

对案例推理技术的研究,从推理过程中不同环节或推理所处阶段来分,包括推理前的处理,如数值缺失处理、模糊值处理、属性约简等;推理中的处理,如索引方法、检索策略等;推理后的处理,如案例的保存、案例库的维护等。

智能决策技术是现代管理中的关键技术,涉及管理科学、人工智能、运筹学和决策科学等诸多领域,是优化理论(包括经典优化理论和现代启发式优化技术)和决策技术的结合。

近年来,基于定性推理的智能决策技术、基于粗糙集的智能决策技术、基于证据理论的智能决策技术、基于群体层次分析的智能决策技术及基于 Multi-Agent 的智能决策技术的研究,极大地促进了智能决策技术的发展^[5~6]。虽然这些技术从不同角度、在不同程度上解决了决策支持系统中的一些问题,但 DSS 仍然难以较好地解决以下问题:领域知识缺乏,专家通过经验解决问题,存

在大量例外于规则的事实,难以建立模型,存在半结构化或非结构化问题,等等。而解决上述问题正是案例推理技术的优势之所在,因此,本文提出了基于案例推理的智能决策技术(Case-Based Reasoning Intelligent Decision Technique, CBR-IDT),来解决这些难题。

案例推理技术是人工智能领域中崛起的一种重要的基于知识的问题求解和学习方法。它解决问题是通过重用或修改以前解决相似问题的方案来实现的。CBR研究方法源自人类的认知心理活动,缓解了常规的知识系统中知识获取的瓶颈问题,它将定量分析与定性分析相结合,具有动态知识库和增量学习的特点。

CBR起源于美国耶鲁大学 Roger Schank 于 1982 年在《Dynamic Memory》中的描述。1995 年召开了第一届国际 CBR 大会,2003 年召开了第五届国际 CBR 大会,每两年召开一次。目前研究重点集中在:案例的索引及检索;案例的修正及其修正规则的获取方法;案例库维护技术及其性能评价;案例工程的自动化;案例推理的理论基础;案例推理与其他方法的集成等。

在 CBR 理论的研究方面,早在 1990 年, Ray Bareiss 和 Bruce Porter 在 PROTOS 系统中就实现了由分类、案例和索引指针构成的网状结构的案例表示^[7]。1994 年, Aamodt 提出了著名的 CBR 四阶段循环,即案例检索、案例复用、案例修正、案例保存^[8],这一结构被广泛认可。1995 年, Gilboa 认为案例可以表示成三元组:〈问题描述,解描述,效果描述〉^[9]。1997 年, Gebhardt 在设计建筑规划项目 FABEL 时发现,很难划分出复杂任务中案例的问题部分和解答部分,因此其在系统推理的动态过程中给予区分^[10]。同年, Aha 认为,对提取案例有困难的应用领域,可以通过“案例工程”来确定案例所必须包含的信息,并从数据中提取信息^[11]。1998 年, Macedo 提出了一种嵌套的图结构案例表示法,引入了层次、空间、时态和因果连接弧,用邻接矩阵表示图结构案例之间的嵌套关系^[12]。1999 年, Andres F. M. Rodriguez 和 Sunil Vadera 提出了基于概率案例的模型 PEBM,其用 Bayesian Network 构造案例的合适描述,并用概率传播技术评估和检索案例^[13]。此外, Leake 在 1999 年研究了如何把 CBR 融合到大规模问题求解任务中,并给出了一些设计原则和方法^[14]。2001 年, CBR 的研究主要集中在案例库维护的各种技术上^[15]。2004 年,我们课题组提出将禁忌搜索(Tabu Search)算法用到案例推理技术中优化 CBR 的搜索过程^[16],同时研究了基于数据挖掘技术的案例推理框架^[17]。2005 年,我们研究了基于离群数据挖掘的案例库维护策略^[18]。2007 年至 2010 年,我们在国家“863”计划项目“面向制造业售后服务商务智能”中研究将案例推理技术用于汽车故障诊

断分析。

在应用开发方面,案例推理技术在规划设计、法律、医学、电子商务及决策支持等领域得到了成功应用。国际上有数个研究机构都在从事该项研究,如美国耶鲁大学的 Roger Schank,佐治亚工学院的 Janet Kolodner,麻省大学的 Edwina Rissland 等,还有德国、英国的研究小组。国内自 20 世纪 90 年代起步后,也有应用成果陆续问世,如 2000 年,中科院智能所研制了基于案例推理技术的农业灾害预报系统^[19];2001 年,中科院计算技术研究所研制了基于案例推理的中心渔场预报专家系统^[20];2004 年,中国农业大学研制了基于案例推理技术的鱼病诊疗系统^[21]等。

1.3 研究意义和内容

1.3.1 研究意义

CBR 技术擅长于解决领域知识缺乏、模型难以建立、存在大量例外于规则的事实的这类半结构化或非结构化问题,而 DSS 则适用于解决问题定义良好、数学模型明确、可归纳出大量规则的这类问题。本书提出用案例推理技术来辅助 DSS 的建造,以充分发挥 CBR 技术和 DSS 各自的优势,弥补其不足,同时引入优化技术、机器学习、知识发现和数据挖掘等技术来深化对 CBR 技术本身的研究。书中利用国际机器学习数据库 UCI 作 Benchmark 分析,利用安徽新安医学数据库数据作实际研究,将提出的 CBR 技术用于其中,探索一种高效、优化、有启发式知识引导的智能决策技术。

目前,在国内外学术界,还没有发现对案例推理智能决策技术 CBR-IDT 系统性的研究,而这项研究对提高 DSS 解决非结构化问题的能力,扩展其应用范围,对 CBR 技术中案例高效的检索和寻优重用都起着关键性的作用。针对近年来国际上兴起的对 CBR 技术的理论和应用研究,对该课题的研究,将有利于促使其在气象、农业、医疗和故障诊断、电子商务等领域得到较好应用,对拓宽计算智能、预测技术的理论研究及研究范围也有一定的意义。

1.3.2 主要研究内容和结构安排

本书主要研究一种基于案例推理的智能决策技术。用案例推理 CBR 技术来辅助决策支持系统的建造,以充分发挥 CBR 技术和 DSS 各自的优势,弥补其不足,同时引入优化技术、机器学习技术来深化对 CBR 技术本身的研究。研究

以 CBR 作为 DSS 中知识表示和推理的核心支持技术;探索 Case-DSS 的系统结构和功能;提出了一种基于 CBR 的集成型、开放式的智能决策技术框架系统;提出了以多策略相似性检索技术作为框架系统的核心支撑技术;研究如何提高 CDSS 的效率和性能。

本书主要包括以下几个方面的内容:

第 1 章分析了研究背景,即决策和智能决策系统、优化技术与优化算法、智能优化算法与启发式算法的研究概况,案例推理技术的研究进展、研究意义和研究的主要内容。

第 2 章给出了一种基于案例推理的智能决策技术 CBR-IDT,并将其应用于构建基于 CBR-IDT 的智能决策支持系统中。在探讨案例推理技术起源、发展、基本思想和特点的基础上,研究了案例推理技术用于决策过程的各个阶段,主要包括:决策案例的表示形式、决策案例的索引、决策案例的组织与检索、决策案例的修正、决策案例的学习和归纳、决策案例库的维护。探讨了案例推理中的核心技术 K-NN 在案例推理中的应用。

CBR-IDT 的智能决策支持系统框架采用人机协调技术,集人的智能和机器的智能于一体,并采用基于案例推理和规则推理相结合的推理方法,提高了对决策过程的智能辅助程度和系统的灵活性、适应性。

第 3 章研究了 CBR-IDT 中属性的优化选择,即基于熵的属性优化选择、基于遗传算法的属性优化选择、基于主成分分析的属性优化选择。在回顾属性选择的策略,包括搜索策略和评价策略,以及属性选择的方法的基础上,对属性选择过程予以形式化。具体研究了基于熵的两种属性选择策略,即信息增益法和增益比率法,用层次化 K-Fold 交叉验证和 K-近邻(K-NN)相结合的技术考察其性能。

利用遗传算法特有的遗传算子搜索机制,用基于相关性的启发式作为评价机制,提出一种 GA-CFS 方法,用于从属性集中选择对给定案例最优的属性子集。对选择出来的属性子集用 C4.5 算法和 K-Fold 交叉验证相结合评价其分类性能。实验对比结果表明,GA-CFS 方法可以确定出与分类和预测最相关的属性子集,同时在几乎不降低分类准确性的情况下,极大地减小属性的表示空间。

研究了基于主成分分析的属性选择,从总体主成分到标准化变量主成分,最后给出样本主成分的计算方法。使用 Iris 数据作为测试,对选出的主成分使用 C4.5 和 K-NN 方法考察其分类准确性。结果表明,使用这种方法可以降低数据维度,同时可以达到一定的分类准确率,当然提取出的主成分的实际意义需要根据具体的问题结合专业知识予以解释。