

中医人体健康状态 机器辨识理论及方法

栾景民 刘旭龙 著

东北大学出版社

· 沈 阳 ·

前 言

中医有着数千年的历史，蕴含着丰富的哲学理论，然而到了近代，由于受到西方医学的冲击，传统中医一度踌躇不前，甚至被全盘否定。究其原因，仅凭四气五味、阴阳五行这种玄妙的理论，无法让世人认识与接受。所以，有必要将传统与现代结合，使中医诊断实现现代化、客观化，由功能想象的描述向阐明机制的方向过渡。马克思说过：“一种科学只有成功地应用数学时，才算达到真正完善的地步。”中医学要成为真正意义上的科学，而不是所谓的经验科学，就必须有数学的参与。西方科学的特点是把数学引进了科学，这促进了西方近代科学技术的革命。现代数学的特点是清晰性、准确性，定量和微观分析是它的长处，这正好解决了中医学描述模糊、不确定及主观性较强、不易把握等问题。

针对辨证诊断过程的复杂性和模糊性，本书以数学形式化高度抽象地表达了中医诊断原理，构建了符合中医“病-证-症”关系的数量化辨证诊断复杂系统模型，提出了一种基于属性偏序结构图表示原理的人体健康状态辨识方法，设计并完成了基于属性偏序结构原理的中医数量化辨证诊断辅助系统原型，开展了临床实验与评估工作，无论对中医诊断客观化的理论研究还是临床的实际应用，均具有重要意义。

全书内容共分7章。第1章为绪论，介绍了人体健康状态辨识的研究背景、需求及研究现状。第2章介绍了属性偏序结构理论的基本思想及成图算法。在偏序理论和形式概念分析的基础上，给出了描述事物之间关联关系的12种属性特征定义，研究了形式背景中属性的属性性质，提出了一种基于属性偏序结构表示原理的知识发现方

法。第3章介绍了基于中医问诊的数量化辨证诊断系统设计及实现。在研究中医辨证中“病-证-症”复杂网络关系的基础上,构建了中医问诊数量化辨证诊断系统模型,并以此为基础,完成了中医问诊数量化辨证诊断辅助系统的设计。第4章在得到的数量化证素集合的基础上,在属性偏序结构理论的框架下,提出了一种基于属性偏序结构图表示的中医诊断证名生成方法。第5章基于本书提出的证素辨识方法,开展了机器辨识系统与名中医专家诊断的双盲对比实验研究,并对实验结果进行了信度和效度评估。第6章介绍了中医诊断中的三支决策问题与中医相容关系证素的辨识方法。通过引入三支决策数量理论,构建了双阈值评价函数的决策风险模型和误判代价评估模型,依据最优决策规则和最小误判代价函数做出最优的诊断决策,为解决中医诊断中多种混合状态下的证素诊断提供了途径。第7章以偏序结构理论为核心,研究了大数据环境下的中医药数据分析及知识发现。根据属性偏序图和对象偏序图中的结构、支路、节点等关联关系,分析了张锡纯治疗中风的32方及用药规律,针对李赛美教授治疗消渴病临床医案的诊疗模式进行了用药规律挖掘,并且对《伤寒论》中对于咳嗽病治疗的方证规律和用药规律进行了知识发现。

本书的出版得到了国家自然科学基金 NO.61401080 和 NO.61601106 的资助,在此表示感谢。感谢广州中医药大学李赛美教授和徐笋晶老师对本书中的医学基础及临床实验提供的指导和帮助;燕山大学的张涛老师和宋佳霖老师对本书提出了宝贵意见,在此深表感谢。

由于作者水平有限,书中难免有疏漏之处,敬请读者批评指正。

著者

2016年11月

目 录

第1章 绪 论	001
1.1 课题的研究背景与社会需求	001
1.2 课题的研究现状	003
1.2.1 中医机器诊断系统的研究现状	003
1.2.2 中医诊断知识发现的研究现状	005
1.2.3 形式概念分析的研究现状	007
1.3 本书的主要研究内容	009
第2章 属性偏序结构理论	012
2.1 引 言	012
2.2 偏序理论基础	012
2.2.1 偏序关系与偏序集	012
2.2.2 形式概念分析	014
2.3 属性特征的定义及性质	017
2.3.1 一般形式背景	017
2.3.2 决策形式背景	024
2.4 属性偏序结构图生成方法	026
2.4.1 属性偏序结构图生成原理	026
2.4.2 属性偏序结构图生成算法	027
2.4.3 属性偏序结构图与概念格	031
2.5 基于属性偏序结构图表示的知识发现方法	033
2.5.1 工程问题上的规则提取	033
2.5.2 医学问题上的知识发现	036
2.6 对象偏序结构图与属性偏序图对比	039
2.7 本章小结	042

第3章 中医数量化辨证诊断辅助系统设计	043
3.1 引言	043
3.2 中医数量化诊断模型	044
3.2.1 病-症-证模型	044
3.2.2 证候-证素-证名模型	045
3.2.3 中医问诊数量化诊断模型	046
3.3 诊断系统的设计	047
3.3.1 系统信息采集	048
3.3.2 系统信息处理	051
3.3.3 中医人体证素状态可视化表示	054
3.4 诊断系统的实现	056
3.5 本章小结	058
第4章 中医辨证诊断证名生成方法研究	060
4.1 引言	060
4.2 证名-证素诊断知识库的构建	061
4.2.1 单病位证素间的结构关系	062
4.2.2 双病位证素间的结构关系	066
4.2.3 无特定病位的病性证素结构关系	068
4.3 中医辨证诊断证名生成算法过程与证名生成结果评估	070
4.3.1 病位诊断知识库的选择	071
4.3.2 属性节点权值的分配	072
4.3.3 中医诊断证名生成方法	074
4.3.4 证名生成结果评估	077
4.4 本章小结	080
第5章 数量化辨证诊断系统临床实验与评估	081
5.1 引言	081
5.2 实验对象与方法	081
5.2.1 实验对象	082
5.2.2 病例纳入与排除标准	082

5.2.3 实验方法	082
5.3 实验数据分析	083
5.3.1 诊断数据的筛选与纳入	083
5.3.2 辨证诊断系统诊断结果与名中医诊断结果的对比	084
5.4 实验结果统计与分析	088
5.4.1 实验结果统计	088
5.4.2 实验结果分析	089
5.4.3 证素及证素组合的频次统计分析	092
5.5 与同类研究的比较	096
5.5.1 中医体质辨识研究	096
5.5.2 中医证候要素组合规律研究	097
5.6 本章小结	098
第6章 人体健康状态辨识中的三支决策问题	100
6.1 引 言	100
6.2 中医理论中的三分思想与三支问题	101
6.3 中医人体健康状态辨识	103
6.3.1 关于人体健康状态的研究	103
6.3.2 基于三支决策数学理论的决策方法	105
6.4 描述人体健康状态的复杂系统模型与证候-证素集合	106
6.4.1 人体健康状态的复杂系统模型	106
6.4.2 中医人体健康状态的证素集合的确定	107
6.5 证素之间的相容关系与非相容关系	108
6.5.1 证候与证素之间的概念关系	109
6.5.2 证素与证素之间的关联关系	110
6.5.3 相容关系证素的确定及构建相容关系证素数据处理模型	111
6.6 基于三支决策数学方法的相容关系证素诊断决策研究	113
6.6.1 基于决策粗糙集理论的三支决策数学模型	113
6.6.2 基于双向阈值评价函数的决策模型	115
6.6.3 基于双评价函数的中医互不包含证素决策模型	116
6.7 中医互不相容证素的误分类代价分析与最小代价决策	119
6.7.1 二支决策的误分类代价估计	119

6.7.2	基于改进误分类代价模型的互不包含关系证素误诊评估	120
6.8	基于自适应阈值调节的个体与群体诊断方法	123
6.8.1	基于自适应阈值调节的个性化诊断方法	123
6.8.2	大人群证素模式规律发现反馈调节自适应阈值参数	124
6.9	本章小结	126
第7章	大数据视角下的中医健康信息知识发现	127
7.1	引言	127
7.2	中医药数据的大数据特征	128
7.3	中医异构数据同构化表示方法	129
7.4	基于覆盖定义的偏序结构生成方法	131
7.4.1	基础覆盖特征定义	131
7.4.2	基础覆盖特征之间的关系	133
7.4.3	基于覆盖定义的偏序结构生成	134
7.5	基于偏序结构的中医数据知识发现	138
7.5.1	偏序结构图中的群结构关系	139
7.5.2	偏序结构图中的子群结构关系	139
7.5.3	偏序结构图中的支路模式	140
7.5.4	偏序结构图中节点之间的关系	140
7.6	基于偏序结构图的张锡纯中风32方知识方法	141
7.7	基于偏序结构图的名中医医案知识发现	146
7.7.1	病历采集	146
7.7.2	数据的规范化及构建形式背景	146
7.7.3	基于偏序结构图的消渴病临床医案知识发现	147
7.8	基于偏序结构图探析《伤寒论》诊疗咳嗽规律发现	151
7.8.1	经方治疗咳嗽方证规律	151
7.8.2	经方治疗咳嗽用药规律	154
7.9	本章小结	156
	参考文献	157

第1章 绪论

1.1 课题的研究背景与社会需求

随着经济社会的快速发展和人民生活方式的转变，人们对健康的消费需求日益提升。我国已经将医学发展的战略优先从“以治愈疾病为目的的高技术追求”转向以“预防疾病的损伤，维持和促进健康”为目标，并将其列入到《国家中长期科学和技术发展规划纲要（2006—2020）》之中。中医学“治未病”的理念和在构建健康保障服务体系方面的优势，得到前所未有的重视。

2015年，国家明确提出要建立中医健康状态评估方法，形成针对不同健康状态人群的中医健康干预方案或指南（服务包），并将中医健康识别系统、智能中医体检系统列为中医药健康服务相关支撑产业重点项目。

辨证论治是中医学的特色和精华，是中医几千年来防病治病的实践积淀和理论结晶。传统辨证诊断的理论依据和思维方式是司外揣内、见微知著和以常衡变^[1]，这种无须打开人体内环境就能推知其内部情况的宏观辨证论治模式，是通过望、闻、问、切等方式所搜集的疾病的起因、病史、症状、体征以及环境因素等临床资料，运用中医知识进行综合分析，辨明其内在联系和各种疾病之间的相互关系，达到对疾病本质的认识，然后干预治疗。中医辨证的思想千百年来指导着中医的临床实践，甚至在难以准确判断病因的情况下也可以辨证施治，改善机体状况。

然而，中医学由于在客观化、量化、标准化等方面的不足，传承和发展受到了严重的制约。传统的中医辨证依旧是靠医生的视觉、听觉、嗅觉、触觉等感觉器官收集症状和体征资料，而根据人体五官收集的信息所得出的诊断结论往往带有很大的主观性。中医辨证思维带有强烈的不确定性和模糊性，且量化与客观化不足，缺少定性定量结合的综合分析，可重复性较差，严重影响了中医的传承和发展。

20世纪60年代，国内外的一些学者对中医诊断客观化进行了探索式研究，尤其是在脉诊和舌诊方面。但是，由于受到客观条件的限制，这些探索式

的研究大多处于散在状态,研究不深、影响不大^[2]。近十几年来,随着现代医学、物理学、工程学等学科的发展和研究领域的交叉融合,中医诊断的信息化、客观化和数量化研究取得了新的进展,出现了脉诊系统、舌诊系统、望诊系统等一系列能够客观采集疾病症状和体征信息并做出智能诊断的仪器系统^[3]。金观昌等采用对多点脉搏波测量,研制了计算机脉诊辅助系统^[4]。Chang等采用阵列式传感器,设计了多点人体脉搏信号采集和诊断系统^[5]。Chung等在采集人体腕脉信息过程中,提出了脉诊脉冲的标准化方法^[6]。高凯设计了一种基于数字图像处理的中医舌象系统^[7]。Zhuo等分析了舌象诊断过程中图片的失真和色彩的校正^[8]。Bai等分析了中医舌象仪的研究进展和现有舌象仪存在的问题^[9]。

辨证诊断的核心思想是司外揣内和见微知著,而在临床的实际应用中,患者的很大一部分自觉症状如头晕、恶心、怕冷等^[10],只能通过问诊来获取。问诊在四诊中收集的临床资料最多、最广,许多信息非问诊不可得^[11]。此外,问诊是获取患者主诉的唯一途径^[12],通过问诊患者能够表达最不舒服的感觉或体征。许朝霞等综述了中医问诊客观化的研究现状与进展^[13]。张永涛等^[14]提出问诊客观化必须首先症状数量化,进而提出了研制中医数字问诊系统的构想。

针对人体健康状态的客观诊断与辨识的研究,最终目的是应用到设计符合中医理论的机器诊断系统。但是,从目前国内外对中医辨证客观化研究采用的技术现状可以看到,迄今为止国内外已有的中医诊断系统尚存在以下问题^[15]:

① 缺乏系统性,整体研究水平仍然较低,并且离临床实用要求还有很大距离。

② 缺乏对辨证施治原理的描述。尽管采用了推理方法,但是没有从根本上解决中医内在辨证施治原理的描述问题。

③ 实用性差。目前的中医诊断专家系统没有能够充分体现系统的“学习”能力和“适应”能力,还不能适用于千变万化的临床医疗。

④ 忽视“整体观”。仅仅针对具体疾病分散开发一些小系统。

针对上述问题,本书以传统的中医学理论为基础,以数学偏序结构理论为中医诊断知识发现工具,以计算机、信息技术为支撑手段,以数量化辨证诊断系统为目标,设计和实现了一种可数量化的、有效的、重复性好且应用于临床的中医辨证诊断辅助系统,以提高中医师的诊疗水平,减少误诊和漏诊的发生;为中医诊断学的数学化描述推理提供了理论基础,同时为数量化辨证诊断机器系统提供了技术支持,具有重大的学术价值和应用价值。

1.2 课题的研究现状

1.2.1 中医机器诊断系统的研究现状

医疗诊断领域是现代化信息处理技术的前沿领域。计算机科学和信息科学在医学诊断上的应用成果早期体现在医学专家系统的研制上。从20世纪70年代开始,人们着手进行医疗诊断专家系统的研制工作。我国学者涂序彦在1977年主持研制了世界上第一个中医专家系统——“关幼波中医肝病诊断处方专家系统”。进入80年代以后,随着计算机技术在中医领域的应用,一系列的中医专家系统应运而生。秦笃烈等^[16]在专著《中医计算机模拟及专家系统概论》中,选编了13个具有代表性的中医专家系统软件。这个时期的专家系统,其诊断信息处理和经验知识学习环节,大部分是基于规则的决策推理。

近些年来,模式识别、机器学习、数据挖掘等技术的迅速发展,也促使了许多先进的对信息获取、信息加工处理及多信息融合的技术进步,涌现了许多新型的中医专家系统^[17-18]。叶进等利用新兴的神经网络技术,设计人工神经网络算法,实现了对病症推断的自动获取^[19]。徐元景等在原有的TCMCADS系统上从脉诊和舌诊入手,建立了以中医教材为基础的诊疗数据库^[20]。“WF文峰-Ⅲ中医(辅助诊疗系统)”采用面向对象方法设计,从多角度进行辅助诊断治疗处理,对内、妇、儿科等460余种疾病进行辅助诊疗^[21]。李海鲲等采用关联规则方法^[22],构建了基于关系数据库的中医诊断系统;并且将辅助诊断系统和医院管理结合起来,使中医专家的思维过程在较大程度上得到模拟,使其具有一定的医院临床使用价值^[23]。丁宏娟、何建成将计算机技术与中医理论相结合,实现了计算机问诊诊断系统,由医生和患者共同进行临床检测,缩减了医生的工作量并提高了医生的诊疗水平^[24]。任晋滔将机器学习技术应用到问诊系统中,以多标记学习算法,提高了证型模式分类的准确度^[25]。李国正等将多标记学习引入到中医问诊中,建立了冠心病的中医辨证分类系统^[26-27]。Chang等根据我国台湾地区偏头痛患者诊断和治疗用药的病例记录,提出了针对偏头痛患者的辅助诊断治疗系统^[28]。

北京中医药大学的王琦教授等在国家重点基础研究发展计划(“973计划”)的资助下,以中医问诊原理的量表方式,开展了中医体质辨识系统研究^[29-33]。通过检索大量的古代及现代文献,结合流行病学调查和临床实践,提炼出个体差异现象的4个表达特征群,即体质特征可以从形态结构、生理机能、心理特

点、反应状态4个方面表达,以此作为寻找人与人之间异同点的标尺,从而辨识体质类型。研究发现,人体的体质类型分别有平和质、气虚质、阳虚质、阴虚质、痰湿质、湿热质、瘀血质、气郁质、特禀质。根据量表设计的原理,在中医体质理论指导下,按照体质类型概念框架的确立和亚量表的设定、条目的收集和条目库的形成、条目的精选、问题的形成、预调查和评价等过程,编制了中医体质量表。形成了由9种体质的9个亚量表结构构成的包含60个条目、以自填为主的标准化量表。

中医体质辨识系统,采用证候的分级量化,通过量表式问卷将问诊信息采集进行了数量化和标准化,在一定程度上对人群和个体的体质进行量化评价,为体质分类研究提供了一个标准化的工具和方法。在全国的流行病学调查中^[34],也证实了人群中确实存在9种体质类型。然而,得到体质类型的辨识结果,用于调理保健毋庸置疑,但是由于相对粗泛的分类结果,体质辨识结论的临床指导意义不足,临床应用性较差,不能为医生或者患者提供有价值的诊断结论和治疗方案。

朱文锋等经过长期的中医辨证规律研究和辨证定量研究,提出了“证素”的概念和由病位证素及病性证素构成的证素辨证新体系^[35-36]。证素辨证体系所确立的辨证思维模式,就是根据症状等临床信息而识别证素,然后由证素组合而做出证名诊断^[37]。由于证候和证素之间是多值映射关系,即每一证候可能对多种证素具有不同的诊断价值,对于某证素的诊断,各证候所起的作用并不均等,因此应当对证候与证素间关系的强度进行定量刻画^[38]。朱文锋等为了解决证素辨证中诊断权值的确定这个关键问题,在“中医辨证数据库”及频数统计基础上,实行“双层频权剪叉”算法,对证素所见证候的权值进行分配,将各证候对各证素和证型的贡献度进行分配,形成了证候标准化权值及证素标准化权值^[39-41]。

在证素辨证体系的核心框架下,早在2002年,朱文锋就研制了基于证素的中医辅助诊疗系统。经过多年的改进和完善,系统内编制有临床症状1000余种,辨证要素60项,常见证型170种、规范证型1700个、演绎证型5000余个。融汇了中医诊断治疗基本的、常规的知识内容,充分体现了辨证论治的特色^[42-43]。

证素辨证体系的建立,对于提高中医诊疗水平、促进中医学的数量化和现代化,具有重要意义。然而,作为一门新兴学科,证素辨证学部分内容尚待完善,如提高辨证计量权值的精确性、完善症状的主次与轻重定量^[44]、不同系统疾病计量权值判定等问题,尤其是由证素组合成证名等问题。证素辨证学中

从证候到证素的研究已基本成熟,但如何将证素组合成证名却成了盲点^[45]。

1.2.2 中医诊断知识发现的研究现状

知识发现(Knowledge Discovery in Database, KDD)是从大量数据中提取出潜在的、可信的、有效且能被人们理解的模式的高级处理过程;是一个由众多学科,如人工智能、机器学习、模式识别、统计学、计算数学、数据库、知识库、数据可视化等相互交叉、相互融合所形成的一个新的研究领域。数据挖掘(Data Mining, DM)是知识发现的核心步骤,是知识发现采用的特定算法或技术。常用的数据挖掘算法有贝叶斯网络、人工神经网络、决策树、遗传算法、最近邻算法、关联规则、模糊分类、粗糙集和形式概念分析等^[46-47]。

医疗领域的数据由于具有动态性、不完整性、冗余性等特点^[48],成为庞大的、结构复杂的数据资源。数据挖掘技术一提出就被应用到医学领域,且研究范围日益广泛。医学数据挖掘的主要目的是发现海量医学数据中的医学诊断规则和决策模式,从而辅助医生进行疾病诊断。中医诊断中通过望、闻、问、切四诊所获知的众多临床资料中隐藏着大量的信息,且中医病症原始数据的采集方法又使其可能混杂有不规范数据或噪声数据,影响了统计分析的结果,而这是中医病症数据的特点之一^[49]。从这些海量的、模糊的、不完备的信息中挖掘知识、发现规律并提供决策模式,是医学数据挖掘的主要目的之一^[50]。

许多学者根据中医诊断具有复杂性、动态性、主观性的特点,将其看作一个复杂的非线性系统,将数据挖掘方法用于中医诊断的内在规律研究中,并取得了较好的效果。

王学伟等应用贝叶斯网络的相关技术从474例血瘀证临床诊断数据中发现了血瘀证的关键症状,定量计算这些症状对诊断的贡献度,建立了血瘀证的诊断模型^[51]。朱咏华、朱文锋将贝叶斯网络运用于中医辨证系统的研究,初步建立起中医辨证数据库,并通过网络学习,形成中医辨证贝叶斯网络结构及概率表,利用建立的贝叶斯网络中医辨证系统,进行数据计量分析,推理验证了证候与证素间的关系。马健等尝试挖掘中医医案中隐藏的临证经验及用药规律,将改进的贝叶斯网络引入到中医医案学习分析^[52]。

朱小虎运用模糊数学方法对膝关节骨性关节炎的中医证候进行客观量化,初步建立该病的辨证分型模糊数学识别模型和临床疗效模糊数学评价模型,探索膝关节骨性关节炎证型概念的模糊数学理论基础^[53]。谢杰采用对于绝经后骨质疏松症4种证型划分法,运用模糊数学的方法对绝经后骨质疏松症患者进行了辨证分型分级研究,构建了较为实用且有远景的绝经后骨质疏松症证型模

糊模式识别数学量化模型^[54]。

吴昊等借助于粗糙集理论的相关知识挖掘中医诊断学中的证候群知识,利用粗糙集理论对临床病历建立知识表达系统,对知识表达系统进行属性约简,求出属性核,建立最小简化表,并且得出有关中医诊断的几个相应规则^[55]。秦中广等运用粗糙集进行中医类风湿证候的诊断,利用属性约简得到寒湿阻络、湿热阻络、痰瘀阻络、气阴两虚、寒热错杂5种证的必定规则和可能规则^[56]。晏峻峰等应用粗糙集的约简理论建立数学模型,分析了症状与证素之间的依存关系,研究了证素决策规则的获取方法^[57]。孙继佳等根据所收集的中医肝硬化患者临床的相关信息,采用粗糙集方法建立证候决策信息表,并通过计算提取与肝硬化各证型有密切相关的重要症状^[58]。

李建生等根据对传网络原理及改进的BP算法网络原理,对Ⅱ型糖尿病文献数据进行证候矢量量化,反映出证候与症状间的关系、证候与证候间的关系^[59]。秦中广等利用粗糙集技术对人工神经网络进行预处理,将改良后的人工神经网络应用到中医类风湿病分型诊断建模,并在实践中取得了令人满意的效果^[60]。李江平等通过分析中医临床数据的特性,将临床数据分为低层级数据和高层级数据,构建了双层的神经网络模型,将其应用到中医辨证智能分类中^[61]。申鹏程在文献[62]中论述了Voila-Jones的人脸检测算法和基于遗传BP神经网络算法在远程中医诊断系统的应用,并对Voila-Jones算法和遗传神经网络算法进行了Matlab仿真。郑景辉等构建了基于BP神经网络的治疗结果拟合模型,进行中医诊断治疗肝硬化的结果预测和影响因素的敏感度分析^[63]。

赵铁牛等认为,证候的诊断是一个在多种方案中如何进行选择决策的问题,决策树解决的核心问题就是数据分类,根据数据的属性将数据分派到不同的组中,决策树为中医证候诊断提供了有力的工具^[64]。文献[65]提出了一种基于决策树的肝硬化辨证诊断分类方法,取得了很好的分类效果。李治等构建了一种基于C4.5算法的决策树模型,对临床阻塞性肺疾病的辨证分型提供了有价值的参考依据^[66]。史琦等采用C4.5,ADTree,CHAID及QUEST决策树方法进行数据分析,构建能够区分不稳定心绞痛病人血瘀证的诊断模式^[67]。

唐伟等运用关联规则的数据挖掘方法,对234例胃脘痛患者中医宏观辨证与胃镜检查结果进行分析研究,提取了胃脘痛患者的胃镜检查结果与中医辨证之间的关联规则,提高了临床胃脘痛患者的辨证准确性^[68]。王亚强等采用关联规则的方法,通过分析证素与证候、证候与症状的关系,得出症状与症状之间的关联,发现了具有相似或相同意义的症状组团^[69]。钟颖等利用关联规则和决策树相结合的方法,研究了中医胃炎临床诊断信息中的“症状”和“辨

证”之间的潜在关系^[70]。

中医学经过两千多年来无数名家的经验传承,从诊断的方法到治疗的方证,都累积了大量的临床资料。如何从这些大量的临床资料和数据中,发现规则、发现模式、发现知识,使辨证诊断的过程可描述、可计算和可推理,更好地推动中医诊疗水平的发展,使之能够更大地造福于社会,是一件意义重大的事情。

1.2.3 形式概念分析的研究现状

“概念”一词,最初是属于哲学的范畴,被理解为由外延和内涵两个部分所组成的思想单元。基于概念的这一哲学理解,1982年德国数学家Wille教授等对概念进行了形式化描述,提出了形式概念分析理论^[71]。在形式概念分析中,概念的外延被看作属于这个概念的所有对象的集合,而内涵则被认为是所有这些对象所共有的属性的集合。概念格是由所有概念连同它们之间的例化与泛化关系而构成的,从本质上描述了对象和特征之间的关系,其相应的Hasse图可以对数据进行可视化表示。近十几年来,形式概念分析作为一种数学形式化的数据处理工具,无论是在其理论领域还是在应用领域,都得到了极大的发展。

以经典形式概念分析的理论框架为基础,许多学者在该方法的改进和推广上做了大量工作。文献[72]将阈值变量引入到形式概念分析中,从而为提取数据分析过程中实际需要的概念提供了简洁而有效的方法。曲开社等^[73]讨论了利用形式概念分析对粗糙集理论的表示与扩展,证明了粗糙集理论中的划分、上下近似等概念可以在衍生形式背景中进行表示。姚一豫^[74]对粗糙概念和经典形式概念进行了系统的比较研究,将逼近算子引入到形式背景中,推导了面向对象的概念格和面向属性的概念格,并发现粗糙概念格同样可以作为完备格的一种具体表示。胡明涵等^[75]将Zadeh模糊理论和形式概念分析相结合,提出了模糊概念格模型,并研究了相应的性质和建格算法。在研究形式概念分析和信息系统的相互关系的过程中,张国强给出了逼近概念的定义^[76],并以此为基础做了许多有意义的推广^[77]。Bendaoud等则在形式概念分析基础上扩展出关联概念分析的思想,并研究了相应的理论和性质^[78]。Dubois等将形式背景分割为几个独立的子背景,探讨了子背景之间的关系^[79]。Kang等在形式概念分析的理论基础上,提出了不同粒度空间下的模糊粒概念^[80]。Djouadi等讨论了划分算子在形式概念分析中的应用,拓展了形式概念分析中模糊格的理论^[81]。Medina研究了形式背景中对象导向的概念格模型和属性导

向的概念格模型,并提出了一种以这两种模型为基础的形式背景属性约简方法^[82]。Yang等提出了一种决策形式背景上的属性约简和规则提取方法^[83]。文献^[84]在形式概念分析的理论基础上,推导了一种新的粗糙集模型,研究了粗糙概念格与传统概念格之间的关系。Eklund等研究了概念格中概念相似度测量的问题^[85-87]。Belohlavek等提出了一种由有序三元组表示的概念格^[88],并在模糊理论的基础上,讨论了概念格中的模糊限制^[89]。李金海等在决策形式背景中,利用保序映射提出了知识约简的概念^[90-91]。这些新的理论模型极大地提高了经典形式概念分析对数据的处理能力^[92]。

近几年,与形式概念分析及概念格相关的应用研究,在本体研究^[93]、软件工程^[94]、知识发现^[95]、语义分析^[96-97]、信息检索^[98-100]等领域也取得了极大的发展。Choi等通过对本体映射的工具、系统、框架以及相关工作的考察,对多种不同类别的本体映射的特点和应用领域进行了讨论^[101]。Eisenbarth等将形式概念分析用于系统需求识别的计算单元特征定位^[102]。Stumme在TITANIC算法的基础上提出用“冰山概念格”的概念,用来分析大型及特大型数据库^[103]。冰山概念格能够在不损失信息的情况下呈现关联规则的频繁模式,同时实现作为挖掘结果的规则的可视化。Yadav提出了用户中心的高品质信息检索的概念化模型^[104],并以此设计了特殊领域高品质信息检索系统。

在国内,形式概念分析也越来越受到人们的关注,论文的数量逐年增加。智慧来将Ganter和Wille定义的Galois连接函数进行了修正,讨论了异构数据集上的偏序形成,提出了面向异构数据分析的广义概念格模型^[105]。许佳卿等将模糊概念分析用于文本分析,提出了一种支持模糊信息的程序聚类方法^[106]。刘树鹏等针对现有的本体合并方法在进行合并时需要人工参与的问题,提出了一种基于形式概念分析的本体合并方法^[107]。何丹丹讨论了形式概念分析在软件工程中的应用现状^[108]。石光莲等讨论了形式概念分析在Folksonomy中的研究进展^[109]。李想等针对信息系统中信息的不完备问题,提出了形式概念分析与粗糙集理论相结合的知识获取模型^[110]。文献^[111]讨论了基于形式概念分析的文本分类研究,提出了一种粗糙形式概念与近邻分类方法相结合的文本分类模型。檀革苗等将形式概念分析和分类树法应用到网络软件的测试中^[112]。刘旭龙等用形式概念分析的方法,提出了一种面瘫严重程度的客观评价方法^[113]。李锋刚等将形式概念分析理论应用于中医痹病的用药规律挖掘中,讨论了挖掘痹病证型、症状的分布情况以及组方用药规律^[114]。刘超男等将形式概念分析应用于《伤寒论》中的方证规律知识获取^[115]。

在形式概念分析的基础上,洪文学等提出了一种基于一种属性偏序原理的分层

结构图表示方法,并将其应用于中医诊断的模式分类^[116-119]、中药配伍组合规律的知识发现等相关研究中^[120-125]。

“中医现代化战略的首要工程是中医知识的现代化表达。而目前的理论瓶颈就在于缺乏东西方思维的表达工具,基于医学知识和医学思维的复杂性,我们建议用复杂概念网络来表达有关人的各方面的知识。”^[126]形式概念分析的核心数据结构——概念格——是提取规则型知识的一个很好的平台,其概念格节点体现了概念的内涵和外延的统一,适合于从复杂网络中发现规则型知识。中医辨证思维的复杂性、证候与证素之间“多对多”的复杂映射关系,都属于复杂网络系统,这些为形式概念分析及其扩展理论应用于中医诊断中提供了用武之地。

1.3 本书的主要研究内容

本书在国家自然科学基金“中医人体健康状态辨识三支决策问题研究”(项目编号:61074130)和“基于面部红外热像温度分布特征的面神经功能定量评估关键技术研究”(项目编号:61401080)的资助下,研究基于属性偏序结构图发现模式的中医问诊的数量化辨证诊断系统。研究内容关联图如图1-1所示。

本书的主要内容包括四大部分。第一部分,在基于属性偏序原理下,发现属性与属性之间的关系,提出形式背景中12类属性特征的定义,并研究相应的性质。在属性特征定义的基础上,提出一种基于属性特征定义的属性偏序结构图生成方法。第二部分是基于中医问诊的数量化辨证诊断辅助系统设计与实现。在中医问诊原理的基础上,设计量表式的问诊问卷,结合中医诊断学理论,构建证候到证素的诊断知识库,并完成了基于中医问诊的数量化辨证诊断系统原型。在测量得到数量化值序证素集合的基础上,构建了证素与证名之间的诊断知识库,提出了一种基于属性偏序结构图的中医辨证诊断证名生成方法。第三部分,在完成的中医问诊数量化辨证诊断系统原型机的基础上,开展数量化辨证诊断系统与名中医专家诊断的盲对实验,对实验的结果进行了统计和分析,并且对本书提出的证素辨识结果进行了评估。第四部分,在临床实验数据的基础上,进一步分析了相容关系与非相容关系证素,针对相容关系证素在表达人体健康状态时表现出的模糊性和不确定性,引入三支决策数学理论方法,构建了双阈值评价函数的诊断决策风险模型和误判代价评估模型。

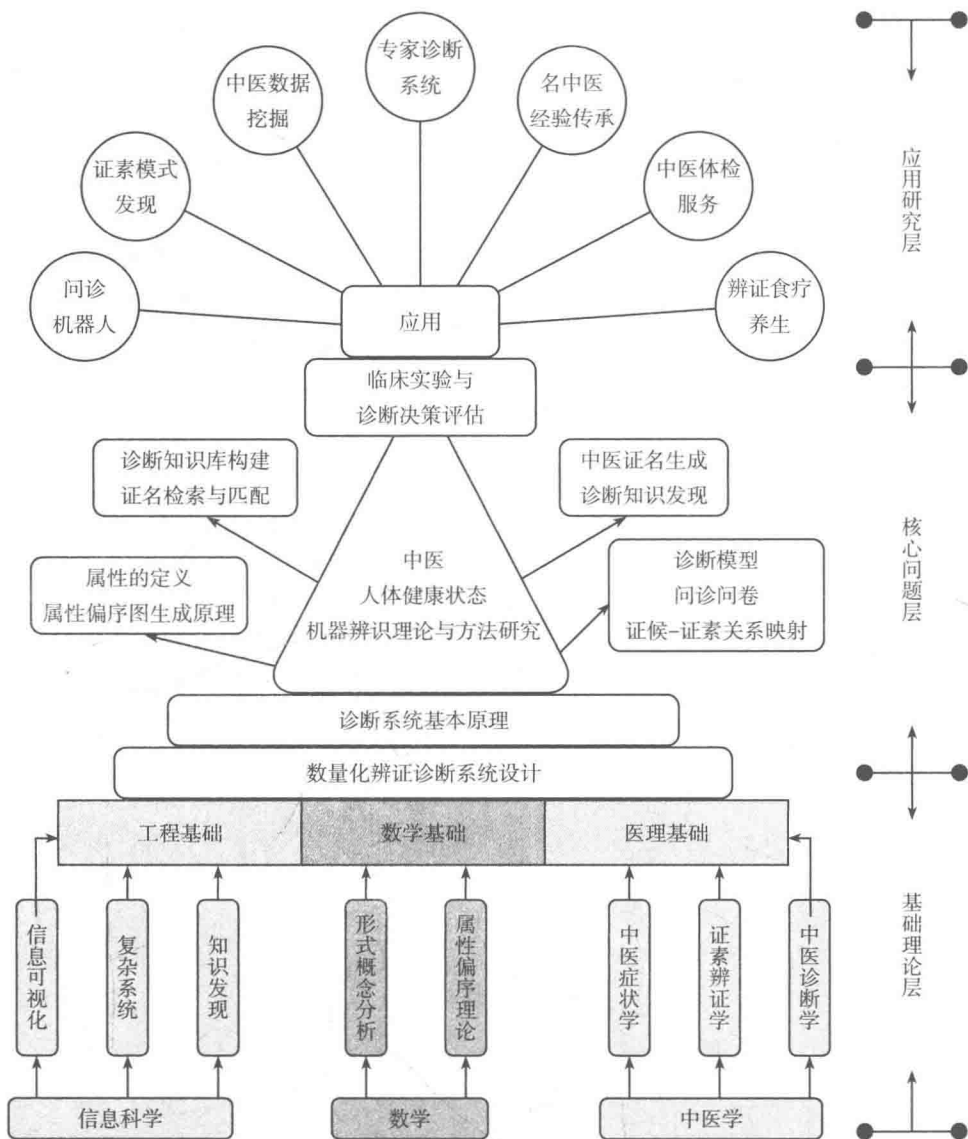


图 1-1 基于属性偏序结构数学原理的中医数量化辨证诊断系统研究内容关联图

本书结构安排如下：

第 1 章，综述中医诊断系统、中医诊断知识发现和形式概念分析的研究现状。

第 2 章，简要叙述偏序理论和形式概念分析的基础，研究形式背景中属性的属性，提出不同特征的完备属性的定义，推导相应的特征属性的性质。在属性特征定义的基础上，提出基于属性特征定义的偏序结构图生成方法。通过国