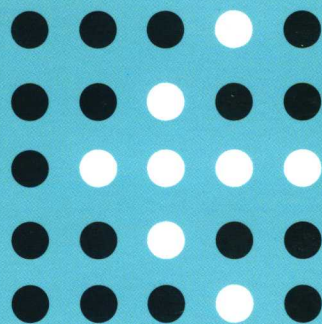




JIYU GUIZE YU ANLI TUILI DE
LINCHUANG JUECE ZHICHI

基于规则与案例推理 的临床决策支持

叶 枫 周根贵 吕旭东◎著



科学出版社

基于规则与案例推理的临床决策支持

叶 枫 周根贵 吕旭东 著

科 学 出 版 社

北 京

内 容 简 介

本书将基于规则推理(RBR)与基于案例推理(CBR)的方法相结合,应用于临床决策支持系统(CDSS)的研究与实践中。基于 RBR 和 CBR 的 CDSS 解决方案,是作者多年来对 CDSS 研究与应用的总结,对 CDSS 的发展以及决策支持系统的发展具有积极的推动作用。本书涉及的部分系统在某三甲医院进行了 CDSS 实验,取得了良好的预期效果。

本书可供管理科学与工程、生物医学工程、计算机科学与技术等专业的高年级本科生、研究生以及相关行业的研究人员阅读和参考。

图书在版编目(CIP)数据

基于规则与案例推理的临床决策支持/叶枫,周根贵,吕旭东著. —北京:科学出版社, 2013

ISBN 978-7-03-038973-2

I. ①基… II. ①叶…②周…③吕… III. ①临床医学—决策支持系统
IV. ①R4-39

中国版本图书馆 CIP 数据核字(2013)第 253439 号

责任编辑:韩卫军/责任校对:宋玲玲
责任印制:邝志强/封面设计:墨创文化

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

四川煤田地质制图印刷厂印刷

科学出版社发行 各地新华书店经销

*

2014 年 1 月第 一 版 开本: B5 (720 × 1000)

2014 年 1 月第一次印刷 印张: 12

字数: 240 000

定价: 59.00 元

(如有印装质量问题,我社负责调换)

**本书获浙江工业大学专著与
研究生教材出版基金资助**

前 言

随着医药卫生事业的发展和科学技术的进步, 医疗行业的科技化水平和信息化程度不断提高。临床决策支持系统(CDSS)作为医学知识工程和人工智能研究中非常活跃的分支, 一直是国内外研究与应用的热点。

临床诊疗的核心是从疾病诊断通过医疗干预有效地控制病情到最终治愈疾病的过程。CDSS 将临床诊疗过程用人工智能的原理和方法来模拟, 包括病人临床数据采集、临床医学知识获取、病人信息与医学知识的匹配、匹配结果的解释以及给出诊疗建议等过程。CDSS 能够通过降低漏诊率、误诊率以及规范化诊疗行为与过程, 达到提高医疗卫生机构医疗服务质量的目的。

但是, 目前已开发的 CDSS 中能够被医生所接受并投入临床使用的还很少。其主要原因在于, 目前的 CDSS 过多地依赖基于临床医学规则知识的推理, 对人类疾病的多样性、多变性和不确定性因素考虑不够。面对错综复杂的病人和疾病情况, 这样的系统起不了太大的作用。基于案例推理(Case-Based Reasoning, CBR)的方法在一定程度上能够解决基于规则推理(Rule-Based Reasoning, RBR)系统的一些问题, 但目前对于 CBR 研究更多的还只是停留在方法层次上。同时, 单独使用基于案例推理的方法也存在一些不足, 例如, CBR 难以表达深层的领域知识, 案例的索引和匹配标准难以统一, 案例检索和相似匹配算法还有待于进一步完善。好在 CBR 是一种方法而非一种技术, 可以不断吸收其他学科的新方法和新技术, 以促进其自身的发展。CBR 与 RBR 的集成能够发挥各自的优势, 从而扩展 CDSS 的应用范围, 改善应用效果。

本书针对目前 CDSS 理论与应用方面存在的一些问题, 通过整合与集成人工智能和计算机技术的相关理论、方法与技术, 探寻 CDSS 发展的新思路, 主要内容包括以下几项。

(1) 基于本体论的思想, 结合对临床医学特点的分析, 给出了构建临床医学本体的方法。在此基础上给出了基于规则推理和基于案例推理的知识表示方法, 包括临床医学知识可视化绘制规范、临床医学知识的 SAGE 建模、临床案例知识表示方法和临床案例知识库的结构。

(2) 知识获取是 CDSS 的关键性问题, 解决好临床医生与知识工程师的沟通问题非常重要。本书分析了临床医学知识获取的特点, 并结合实例给出了知识的流程图化表示、医学知识的 SAGE 建模、规则知识生成和知识获取流程。文本信息抽取是知识获取的第一步, 对于案例知识的获取具有重要作用。本书通过实验

阐述了基于有限状态自动机和基于条件随机场的医学命名实体识别方法，并对实验结果进行了分析，给出了临床案例知识获取的方法和流程。

(3) 本书分析了基于规则推理的正向推理、反向推理和正反向混合推理的逻辑原理，以及基于案例的 CBR 四阶段循环理论，提出了 CBR-RBR 集成推理模式，介绍了基于 SAGE 建模的规则库构建方法和规则推理的主要步骤。相似度计算是案例推理中最关键的问题，尤其是对于具有灵活结构的临床案例来讲。在现有相似度计算方法的基础上，设计了针对临床案例特点的基于文本类型、程度类型、数值类型、时间类型等多种数据类型的相似度计算方法。为了实现相似度权重的自动计算，本书提出了基于 TF-IDF 思想的项目权重计算方法和基于自组织竞争神经网络的综合权重计算方法，并通过基于真实临床数据的实验验证了以上方法的有效性。

(4) 本书提出了基于临床案例相似度矩阵的层次聚类分析方法和基于案例相似度矩阵范数的凝聚度分析方法。通过对 II 型糖尿病案例和高血压案例的分析，论证了该分析方法的意义和作用。同时给出了一个 CDSS 的系统框架，并介绍了系统的主要功能，分析与评价了在某三甲医院进行的临床决策支持系统实验。

本书以 CDSS 的临床应用为导向展开理论与方法的研究，研究过程中强调针对临床医学诊疗过程的特点，充分发挥基于案例推理与基于规则推理的优势。通过增强临床案例知识表示方法的灵活性和适应性，加快临床医学知识库建设的步伐。实践证明，理论与临床实践相结合对 CDSS 的研究与发展至关重要。

叶 枫

2013 年 9 月 20 日于浙江工业大学

目 录

前言

第 1 章 绪论	1
1.1 决策支持系统的概念	1
1.2 CDSS 的概念	4
1.3 CDSS 研究现状	6
1.3.1 CDSS 的发展历程	6
1.3.2 基于临床指南的 CDSS	7
1.3.3 RBR 与 CBR 的 CDSS	8
1.3.4 基于本体的 CDSS	9
第 2 章 相关理论与方法	11
2.1 本体与医学本体	11
2.1.1 本体与领域本体	11
2.1.2 医学本体的抽象化表示	12
2.1.3 医学本体描述语言	14
2.2 临床医学知识表示方法	15
2.2.1 知识的形式化表示方法	15
2.2.2 临床医学知识表示模型	19
2.3 规则知识获取的方法	21
2.3.1 数据库中的知识发现	22
2.3.2 规则知识的获取流程	23
2.4 文本信息抽取方法	24
2.4.1 命名实体识别方法	25
2.4.2 实体关系抽取方法	26
2.4.3 句法分析	27
2.4.4 有限状态自动机与正则语法	28
2.4.5 基于条件随机场的命名实体识别方法	30
2.5 基于规则的推理方法	32
2.5.1 正向推理	32
2.5.2 反向推理	34
2.5.3 正反向混合推理	36

2.5.4	冲突解决策略	37
2.6	基于案例的推理方法	37
2.6.1	相似度计算方法	38
2.6.2	相似度权重的智能计算方法	40
2.6.3	案例相似度矩阵范数与层次聚类分析方法	43
第3章	临床医学知识的表示	47
3.1	医学本体的构建	47
3.1.1	医学本体概念词典的构建	47
3.1.2	医学本体的层次结构	49
3.2	临床医学规则知识的表示	50
3.2.1	临床医学知识可视化表示规范	50
3.2.2	临床医学知识的 SAGE 建模	53
3.3	临床医学案例知识的表示	55
3.3.1	临床案例知识库结构	56
3.3.2	临床案例知识库设计	60
第4章	临床医学知识获取方法	66
4.1	规则知识的获取方法	66
4.2	文本信息的抽取方法	72
4.2.1	文本信息抽取的一般过程	72
4.2.2	医学命名实体的识别流程	73
4.2.3	基于条件随机场的中文命名实体识别	76
4.3	案例知识的获取	81
4.3.1	电子病历	81
4.3.2	案例知识获取流程	82
4.3.3	临床案例结构化信息的获取	83
4.3.4	临床案例知识的 XML 表示	85
第5章	规则推理与案例推理的实现	93
5.1	CBR-RBR 集成推理模式	93
5.2	基于规则推理的方法	96
5.2.1	基于 SAGE 模型的规则库构建方法	96
5.2.2	规则推理的实现	99
5.3	临床案例相似度计算	100
5.3.1	属性值相似度计算方法	101
5.3.2	匹配方式与相似度计算的关系	107
5.3.3	相似度计算中权重的确定	109

5.3.4 基于 TF-IDF 原理的项目权重计算	109
5.3.5 基于自组织竞争神经网络的综合权重计算	115
5.4 临床案例相似度计算综合实验	121
5.4.1 实验数据的采集与预处理	121
5.4.2 临床案例样本之间的相似度计算	122
5.4.3 实验结果分析	141
5.4.4 实验总结	145
5.5 基于临床案例的层次聚类分析	146
5.5.1 基于临床案例知识库的层次聚类分析	146
5.5.2 临床案例聚类分析的意义	151
5.6 基于案例推理的方法	152
5.6.1 临床案例推理的主要过程	152
5.6.2 项目权重和综合权重的保存	155
5.6.3 案例检索	156
第 6 章 基于规则推理与案例推理的 CDSS 实践	158
6.1 CDSS 的设计原则	158
6.2 系统框架	158
6.3 系统主要功能设计	159
6.3.1 规则推理的功能设计	160
6.3.2 案例推理的功能设计	164
6.4 CDSS 临床实验及评估	168
6.4.1 CDSS 临床实验	168
6.4.2 CDSS 临床实验的结果评估	169
参考文献	173
索引	181

DSS 经历了 40 多年的发展历程。20 世纪 70 年代末 DSS 的基本框架形成,明确了数据与模型的结合是 DSS 区别于其他信息系统的主要标志。1978—1988 年, DSS 发展迅速,出现了许多基于 DSS 的系统。这一阶段的 DSS 主要由数据库、模型库和人机交互系统组成,但是多数 DSS 模型化功能较弱,人机接口不够友好,系统对环境的适应性较差,因此,真正投入运行的 DSS 非常少。20 世纪 80 年代末,人工智能与 DSS 开始互相结合,比较典型的是专家系统与 DSS 结合形成的智能决策支持系统(IDSS)。进入 21 世纪后, DSS 开始向深度和广度发展,数据仓库与数据挖掘的兴起大大推动了 DSS 的应用和发展。DSS 在许多行业的研究与应用取得了重要的成果。

除了传统的 DSS 外,还有许多具有不同特征的 DSS^[4],可以概括为以下几种类型。

1. 群体决策支持系统

需要两人或多人一起进行问题的讨论、问题解决方案或策略的提出、方案的评价以及最终作出决策,这样的决策过程称为群体决策。群体决策是决策时间、地点、通信网络、个人偏好以及其他一些技术问题的复杂组合。群体决策支持系统(Group Decision Support System, GDSS)的目的就在于为群体决策过程提供一种有组织的指导信息交流方式、议事日程、讨论形式以及决议过程的系统平台。

GDSS 利用计算机网络与通信技术,能够为身处不同地理位置的多个决策者提供一个具有良好交互沟通功能的决策支持系统,该系统能够帮助决策者通过某种规程以相互协作的方式探寻半结构化或非结构化决策问题。GDSS 与传统的群体决策方式相比,具有以下优势:不受时间与空间的限制;能让决策者相互之间便捷地交流信息与共享信息,减少片面性;能集思广益,激发决策者的思路,使问题的解决方案尽可能趋于完美;可防止小集体主义及个性对决策结果的影响;可提高决策群体成员对决策结果的满意度和置信度。GDSS 的应用研究已经取得令人瞩目的进展,尤其是互联网技术和移动通信技术的快速发展;电子公告板、电话会议、远程视频会议和即时消息系统等的应用,为 GDSS 的应用奠定了基础,并进一步拓展了 GDSS 的发展空间。

2. 分布式决策支持系统

随着全球化经济的不断发展,企业所面临的决策环境越来越复杂,许多大规模的管理决策活动已不可能或不便于采用集中的方式进行。分布式决策支持系统(Distributed Decision Support System, DDSS)针对分布式的决策环境,进一步扩展 GDSS,以支持组织中决策网络节点的决策、通信、协调和合作。

DDSS 的特点是：能支持处于不同节点的多层次的决策，提供个人支持、群体支持和组织支持；基于分布式计算网络平台，尤其是云计算平台出现之后，使得资源共享与运行效率达到比较理想的程度；各节点间提供交流机制，支持人机交互、机机交互和人人交互；具有处理节点间可能发生的冲突的能力，能协调各节点的操作；既有严格的内部协议，又是开放性的，允许系统或节点方便地扩展；系统内的节点作为平等成员，不形成递阶结构，每个节点享有自治权。

3. 智能决策支持系统

智能决策支持系统(Intelligence Decision Support System, IDSS)的概念是在 20 世纪 80 年代由美国学者 Bonczek 等提出的。IDSS 的核心思想是将人工智能以及其他相关学科的成果引入 DSS 中，使 DSS 具有人工智能。DSS 中引入专家系统的主要优势体现在可以处理定性的、近似的或不精确的知识。人工智能在改善 DSS 的人机接口水平，尤其是对话功能上具有重要贡献。

IDSS 一直是 DSS 研究与应用领域最活跃的分支。有关 IDSS 的结构有很多种版本，比较经典的 IDSS 结构包括以下几部分。

(1) 用户接口模块，即 IDSS 的人机交互接口，帮助实现系统与用户之间的有效地沟通。如何提高人机交互的人性化程度，以及采用接近自然语言的方式进行人机对话，一直是 DSS 研究的热点之一。

(2) 数据库系统，主要包括数据库管理系统与数据库，用于保存 IDSS 所必需的原始数据、运行过程的中间数据以及结果数据。在许多系统中，知识库以及模型库也是以数据化的形式保存的，因此，数据库是知识和模型的一种载体。

(3) 知识库系统，一般由知识库管理系统、知识库及推理机三部分组成，用于实现对知识库中知识的增、删、改等知识维护与管理功能。知识库是知识库系统的核心，用于存储决策专家的知识和经验，以及一些特定问题领域的专门知识。知识库中知识表示的方法与知识的推理方式相关，甚至会对推理方式产生影响。推理机是实现推理过程的一组程序，它针对用户问题来处理知识库(规则和事实)。

(4) 模型库系统，包括模型库管理系统(MBMS)和模型库(MB)。模型库系统是传统 DSS 的三大支柱之一，它可以通过与决策者的交互构造符合实际的、合适的模型。MBMS 主要实现模型的生成、修改、更新、删除、连接及运行控制等功能。在许多模型库系统中，模型需要经常与数据库交互，因此，模型库系统与数据库系统之间存在许多关联的功能。

4. 基于数据仓库的 DSS

20 世纪 90 年代发展起来的数据仓库和数据挖掘技术,为解决传统 DSS 存在的一些问题提供了技术支持,使 DSS 发展进入了一个新的阶段。

数据仓库是面向主题的、集成的、不可更新的、随时间变化的数据集合,用以支持管理活动中的决策制定过程。数据仓库以多维数据为核心,能够满足 DSS 的数据分析要求,克服了传统数据库管理系统的数据组织性差、利用率低的缺点。其中联机分析处理(OLAP)是基于数据仓库的信息分析处理过程,专门用于支持复杂的分析操作,侧重对决策人员和高层管理人员的决策支持,可以根据分析人员的要求快速灵活地进行大数据量的复杂查询处理,并且以一种直观易懂的形式将查询结果提供给决策人员。数据挖掘则是利用各种软件工具在大量的数据中找出隐藏的模式或关系,并从中推出规则,用于预测未来的行为以及作为决策的依据。

基于数据仓库的 DSS 一般以数据仓库为核心,以 OLAP 和数据挖掘技术为主体,为决策者提供集成的信息查询与报表功能,以及面向问题的数据建模、数据挖掘与分析功能。数据仓库与数据挖掘技术还可应用于 IDSS、GDSS 和 DDSS 等 DSS 中,进一步扩展 DSS 的研究与应用领域。

1.2 CDSS 的概念

临床医生给病人看病的过程是一个对病人信息进行综合分析处理的自然过程,一般可将这个过程分为 3 个阶段^[5]:观察(observation)阶段、诊断(diagnose)阶段和治疗(therapy)阶段,诊断阶段和治疗阶段可以回到前面的阶段,该循环过程如图 1-1 所示。这个过程包括病人讲述病史、临床症状、疾病表现和原因以及

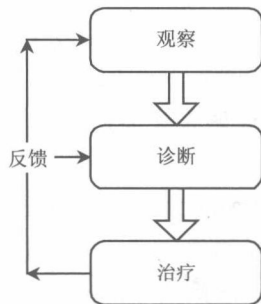


图 1-1 医生的诊疗循环

病患家族史等情况,临床医生采集相关诊疗数据(如物理检查、实验室检查等),回忆或搜索相关医学知识和临床经验,针对病人的实际情况进行比较或类比,并得出结论或给出诊断结果,最后为病人开出治疗处方或执行其他治疗措施。从系统的角度来看,观察阶段的主要任务是获取与诊断相关的信息,对病人相关信息理解和挖掘得越充分,对患病情况的了解就越多。医生根据病人描述的信息以及相关检查检验信息,结合其所掌握的知识

和经验,给出诊断结论的过程实质上是一个推理过程,而治疗则是通过求解和处理问题实现对病情的有效控制,并最终治愈疾病的过程。治疗依赖于诊断和决策分析,而医生诊断和决策分析的准确性与病人信息的获取

程度以及医生的知识和经验密切相关。

医生给病人看病的过程用人工智能的原理和方法来模拟,这就是 CDSS。它是 IDSS 在临床医学辅助决策领域的一个重要分支。CDSS 对医生看病的观察、诊断和治疗的模拟过程包括病人数据收集、医生临床医学知识和经验的收集、病人信息与医学规则的匹配、匹配结果的解释以及医生给出诊疗建议等过程^[5]。

Musen (1997) 对 CDSS 作了如下定义^[6]: 任何将临床数据作为输入信息,将推论结果作为输出信息,有助于临床医生决策并被用户认为具有智能性的软件即称为 CDSS。

CDSS 就是通过计算机为临床医生的决策提供辅助工具的^[7]。CDSS 是医学知识工程和人工智能研究中非常活跃的分支,目前大多数的 CDSS 是运用专家系统的设计原理与方法,利用医学专家的知识和治疗疾病的思维过程建立的计算机系统,其基本组成如图 1-2 所示。

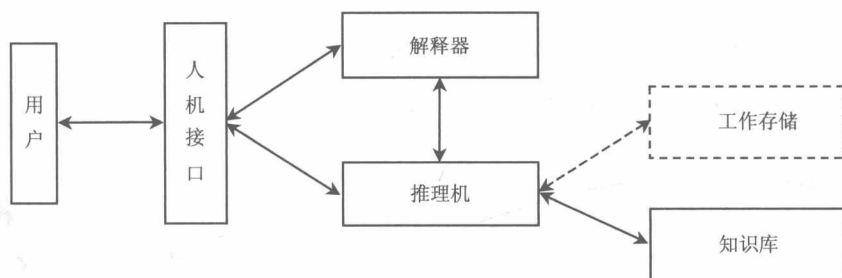


图 1-2 基于专家系统的 CDSS 的组成^[8]

推理机是知识规则应用于问题求解的载体。它负责控制、协调整个专家系统,并根据输入的数据,利用知识库(knowledge base, KB)中的知识(规则),按照一定的推理策略进行问题的求解。推理引擎使用的知识在知识库中表达。

知识库是知识工程中的知识集群,具有领域特征,基于若干知识表示方式存储于计算机中的知识片集合。这些知识片包括与领域相关的知识、事实数据,以及由专家经验得到的启发式知识等。知识库的建立一般需要通过相关领域的专家与知识工程师合作来完成,部分知识(如一些事实数据)可以通过计算机自动获取。

工作存储(working memory)即数据中心,主要指存放于数据库或其他存储形式与临床决策相关的信息,如患者基本信息、病案史、家族史、个人史等,还包括系统运行中用户输入的一些中间数据、推理得到的中间结果及最终结果等。

解释器负责向用户解释根据患者信息运用医学知识推理的过程和结论。

人机接口(User Interface, UI)用于实现系统与用户的对话。用户通过人机接口输入数据、提出问题并获得系统的推理结果,系统则通过人机接口向用户提出询问。

1.3 CDSS 研究现状

1.3.1 CDSS 的发展历程

世界上第一个功能较全面的 CDSS 是 20 世纪 70 年代斯坦福大学的 Shortliffe 等研发的 MYCIN^[9], 该系统主要用于中枢神经系统感染的诊断和治疗咨询, 用户输入患者的临床表现和化验诊断结果, 系统可以据此判断感染细菌的种类, 并提出临床处理的建议。MYCIN 定位于医生顾问, 具有一定的解释功能, 它对医疗领域乃至整个专家系统的发展具有重大影响。1982 年, 匹兹堡大学的 Miller 等研制了 Internist-I 内科计算机辅助诊断系统, 其知识库包含 572 种疾病和 4500 多种症状^[10]。1991 年, 哈佛医学院开发的 DXPLAIN 系统包含 2200 种疾病和 5000 多种症状^[11]。除了能诊断多种疾病的大型专家系统外, 针对某一种疾病的专项医学专家系统更是举不胜举。例如, 1996 年 Birndorf 等开发的贫血诊断报告系统^[12], 2000 年 Wells 等开发的计算机辅助乳腺治疗计划系统^[13]等。人工智能应用于 CDSS 的研究一直是国内外研究的热点之一, 专家系统、神经网络、机器学习以及一些优化学习算法与 CDSS 的结合产生了一批研究成果^[14-16]。

20 世纪 70 年代我国就开始了 CDSS 的研究。1975—1979 年, 北京中医医院、中国科学院自动化研究所、北京第二医学院合作研究编制的中医关幼波教授对肝炎的辨证论治诊疗经验的电子计算机程序, 是我国最早基于中医理论的 CDSS。它采用模糊逻辑算法表达了中医辨证施治的专家知识。20 世纪 80 年代以来, 国内研制的中医诊疗专家系统有 100 多个, 基本上都是基于规则的决策推理。例如, 罗元恺痛经辨证施治系统采用树型结构的知识库; 钱伯煊月经病诊疗经验模拟则是利用通用工具中医诊疗生成系统开发的, 其规则是产生式专家系统结构。这些系统受到早期专家系统的限制, 问答式切入过程太长, 专家系统的知识面相对较窄, 医生的思维及病人病情只能跟着专家系统路径走, 很容易超出系统边界, 而得不到圆满的诊断结果, 所以至今几乎没有一个能保持临床应用。另外, 有一批专科 CDSS, 如贫血鉴别系统、肝病营养疗法专家系统、急性肾衰诊断系统、甲亢诊断系统、颈腰疾病专家系统、肺癌诊断系统、心功能辅助诊断系统等, 其中部分系统在小范围内得到了应用。

大多数人都认同 CDSS 具有改善临床诊疗, 至少有改变医生行为的作用^[17-18]。CDSS 在提醒、警告、制订治疗计划以及患者的教导方面都能有效地改善临床实践行为^[19]。CDSS 给予临床医生的决策支持方式不是单一的, 支持临床决策的方式可以分为主动式和被动式两种。主动式就是无论当时医生主观上是否需要决策支持, CDSS 都主动给出决策建议, 以强制性地阻止一些严重后果的发生, 例如,

用药配伍禁忌、药物-疾病禁忌等。这种方式的临床决策支持已经被融入一些电子病历或医院信息系统(HIS)中。所谓被动式决策支持,是指医生通过与CDSS交互发起决策支持请求,CDSS给出临床诊疗建议的方式。

1.3.2 基于临床指南的CDSS

临床指南是系统开发的多组临床指导意见,用于帮助医生和病人针对特定的临床问题作出恰当的处理和选择,决策适宜的卫生保健服务^[20]。临床指南在广泛收集临床证据的基础上,基于循证医学(Evidence-Based Medicine, EBM)的方法进行开发,它的开发方法严格,内容科学准确,是指导医疗研究和实践的重要手段,它的目的在于提高临床诊疗的安全性、质量以及成本效益^[21]。1990年美国医学研究所发表报告以来,临床指南的重要性得到了广泛认可^[22]。NGC(National Guideline Clearinghouse, 国家临床指南中心)是美国卫生保健研究和质量局建立的,面向大众的免费临床实践指南数据库,截至2006年5月,其下已存有1900余篇指南全文^[23]。

作为CDSS知识库重要组成部分的临床指南,起着向推理机提供临床诊疗知识的重要作用^[24-25]。然而,大多数的临床指南都是文本形式的,将其转化为计算机可识别的临床指南(Computer-Interpretable-Guideline, CIG)非常关键。近年来,世界各国都在进行着不同程度、不同方式的临床指南计算机化研究工作。在国际上公开发表的比较成熟的指南表达模型有近20种^[24],而国内在此研究领域还没有标志性的研究成果。

Arden Syntax临床指南模型于1990年开始发布,后被HL7(Health Level 7)和ANSI(American National Standards Institute, 美国国家标准学会)采纳为标准,是最早的医学知识表达标准。Arden Syntax临床指南模型结构简单,提供了数据引用、触发事件和消息机制,但它只适合单步决策,且没有明确的定义提醒机制^[26-27]。著名的HELP(Health Evaluation through Logical Processing, 经逻辑处理的健康评估)系统是较早的基于临床指南的CDSS。在HELP系统中,采用标准的Arden语法描述医学逻辑模块、基于HL7消息机制整合于各个临床信息子系统中^[28]。

GLARE(Guideline Acquisition, Representation and Execution)自定义了一套基于图形的表达模型对临床指南进行编码,并和推理引擎分开,形成指南获取、指南执行两个相对独立的模块^[29]。GLIF临床指南模型是在Arden Syntax临床指南模型的基础上发展起来,同时吸取了Asbru、GUIDE、PRODIGY、PROforma、EON等优点,它的医学数据模型基于HL7 RIM,其面向对象的查询表达语言GELLO被HL7采纳为标准^[30-31]。2002年发布的SAGE临床指南模型综合了以往的研究成果,基于HL7 RIM以及SNOMED临床术语等发展标准,采用了事件驱动方式^[32]。

基于临床指南的 CDSS, 将静态文本形式的临床指南按照合适的指南表达模型进行建模, 形成计算机可理解的模型, 并能够结合患者信息执行指南, 在诊疗的关键时刻提供针对性建议。以这样的方式利用临床指南比查阅在专题论文或学报里出版的临床指南, 或者利用那些通用的、与患者的个体信息无关的临床指南要方便有效得多。研究开发基于临床指南的 CDSS 有助于提高医生的诊疗水平, 提高医疗服务质量, 最终使患者受益。

1.3.3 RBR 与 CBR 的 CDSS

RBR 的系统又称为产生式系统, 其表现形式单一且易于用户理解, 因此成为最重要的知识表现方法。目前, 研究和应用的 CDSS 中 RBR 的占绝大多数^[33]。但这种推理方式存在一些明显的不足, 其中最主要的问题是规则知识的获取十分困难, 系统的自学习功能非常弱。另外, 规则间的相互关系表达不清楚, 知识的整体形象难以把握, 容易出现控制饱和问题等。但这并不能说明这种方法就没有发展的空间了, 国内外在 RBR 方面的研究从未停止过^[34-35], 在一些特定的领域, RBR 还是能够发挥其重要作用的。

CBR 是人工智能领域一种重要的推理方法, 它通过与以往案例知识的检索与匹配给出新问题的求解过程和结果。CBR 对领域知识模型的要求相对宽松许多, 大大方便了领域知识的获取。用户可以自己增加、修改和完善案例, 增量式的学习使案例库的覆盖度随系统的使用逐渐增大, 系统的判断能力不断提高。这种方法在一定程度上能够解决传统推理方法的一些固有问题。因此, 在 CDSS 领域展开了广泛的研究^[36-39]。但是, 单独使用 CBR 也存在一定的不足, 如难以表达深层的领域知识, 庞大的案例库对于存储计算平衡等问题是个挑战, 案例的索引和匹配标准难以统一等。

相对于其他人工智能技术, CBR 在 CDSS 中的应用具有一定的优势^[39-40], 主要体现在医生诊疗过程所依赖的经验主要由其记忆中典型的、特殊的病例构成, CBR 可以很好地描述医生的这种经验。另外, 在医生诊断的过程中, 特别是在提出诊断假设的过程中, 往往会回忆与当前病情相似的病例, 参考它们的治疗方案, 根据案例与当前病情的差异, 在先前治疗方案的基础上作出相应改动, 新病例还可以作为一个新案例被保存。

CBR 和 RBR 各有优势。RBR 较 CBR 而言, 占用空间要少得多, 而且推理过程严谨, 能够很快给出明确的结论。但规则是有限的, 规则的形成需要一个比较复杂的过程, 往往会有些滞后, 例如, 临床指南作为制定规则的一个主要依据, 有滞后期。疾病每天都在不断地增加和变化, 大多数疾病还没有形成明确的推理规则。作为 RBR 方法的重要补充, CBR 与 RBR 相结合的方法能够弥补

一定的空缺,国内外在这方面的研究文献比较多^[41-44]。

1.3.4 基于本体的 CDSS

本体(ontology)一词是由17世纪的德国经济学家 Goclenius(1547—1628年)首先使用的,它原是哲学中的概念,研究客观事物存在的本质。本体的含义是客观存在并形成表象的根本实体。

在人工智能界,1991年,Neches等最早给出了本体的定义:给出构成相关领域词汇的基本术语和关系,以及利用这些术语和关系构成的规定这些词汇外延规则的^[45]。1993年,Gruber将本体定义为“概念模型的明确的规范说明”^[46]。

人类知识基于概念与概念之间的相互关系进行表述,而人类组织知识的方法则是基于概念的分类,这是基于工程的本体论的主要观点。本体论的应用范围越来越广,在很多有名的知识系统中都不同程度地应用了本体论。如美国的 Lenat 领导的小组研制的常识知识库系统 CYC,普林斯顿大学伯克利分校研制的语言知识库 WordNet^[47],以及陆汝铃等研制的常识知识系统^[48]等。

基于本体的知识表示方法为医学领域知识的应用打下了坚实的基础。利用本体构建的知识库具有统一的概念术语规范,对知识获取、分析和共享很有帮助。基于本体从病例中挖掘信息,辅助诊断治疗的方法,与传统数据挖掘领域或机器学习领域中分类法的区别在于:首先通过建立医学领域本体形成相关先验知识,之后对病例信息进行归纳学习,而不是像传统分类法中知识的表示方式那样,根据具体问题而定,各不相同,难以在不同的应用之间共享知识,不具备伸缩性。

医学知识是人类最重要的知识之一,世界各国都十分重视医学知识工程方面的研究。美国国立医学图书馆著名的统一医学语言系统(Unified Medical Language System, UMLS)已经成为医学领域的医学语言表示标准,为生物医学领域相关知识的应用提供了相对统一、可以共享的信息资源^[49]。UMLS 有两个最主要特征^[50]:①规范医学用语,将概念的不同表达方式进行了同义规范;②通过分析概念的内涵和外延,建立了概念间的上下位(分属)及相关关系。虽然 UMLS 建立了一个医学本体术语体系,并描述了部分术语间的关系,但是没有利用这些术语描述医学事实。世界卫生组织及各成员国相互协作完成的国际疾病分类——ICD-10^[51],采用多轴心的分类系统将疾病纳入相应的类目下,为每种疾病提供了一种分类学上的类目,为疾病分类提供了一种国际参考标准。

需要指出的是,医学本体体系并不只是一种描述语言,而是独立于任何语言之上的客观存在。对医学这样一个巨大而复杂的系统来说,其本体的结构也十分复杂,要给出比较完整全面的描述是十分困难的。