

文章编号: 1671-7104(2009)01-027-04

## 基于SAGE的医学知识模型的设计与应用

【作者】 杨艳<sup>1</sup>, 吴彬飞<sup>2</sup>, 叶枫<sup>1\*</sup>, 吕旭东<sup>2</sup>

1 浙江工业大学经贸管理学院 (浙江, 杭州, 310023)

2 浙江大学生物医学工程教育部重点实验室 (浙江, 杭州, 310027)

【摘要】 本文分析了SAGE这一医学知识模型, 并以代谢综合征诊疗知识为例阐述SAGE建模方法。针对SAGE表达能力的不足, 提出了改进方法。将模型构建为知识库在CDSS中进行应用, CDSS的评估结果表明, SAGE模型具有临床应用价值。最后还提出了SAGE模型仍需继续改进之处。

【关键词】 SAGE; 知识库; 建模; 临床决策支持系统

【中图分类号】 TP311.132

【文献标识码】 B

## Design and Application of Medical Knowledge Model on SAGE

【Writers】 YANG Yan<sup>1</sup>, WU Bin-fei<sup>2</sup>, YE Feng<sup>1\*</sup>, LV Xu-dong<sup>2</sup>

1 College of Business Administration, Zhejiang University of Technology, Hangzhou, 310023

2 The key Laboratory of Biomedical Engineering in Ministry of Education, Zhejiang University, Hangzhou, 310027

【Abstract】 As an methodology for promoting the quality and efficiency of health care, clinical decision support systems(CDSSs) have gained much improvement. The knowledge base(KB)plays an important role in DSS. For CDSSs, the construction of KB means modeling the medical knowledge based on a suitable model. This study analyzes the SAGE model, then implements it on knowledge of diagnosis and treatment of Metabolic Syndrome(MS), and improves the SAGE to enhance its expression ability. The model is constructed as the KB in CDSS, and be applied in hospital. The evaluation result of CDSS reveals that the SAGE model should be useful in clinical application. Finally, this study propounds some points yet to be improved in the SAGE.

【Key words】 SAGE(Standards-Based Sharable Active Guideline Environment), Knowledge Base(KB), Modeling, Clinical Decision Support System(CDSS).

如何提高国民健康水平, 提供高质量低负担的医疗服务, 是世界上许多国家日益关注的问题。解决这个问题的途径之一便是利用信息技术服务于医疗领域, 提高医疗质量和效率, 减少医疗差错, 降低医疗费用。临床医生对信息技术的需求是必然的, 当他们为患者进行诊断和制定防治计划时, 常常需要计算机的辅助。因为<sup>[1]</sup>: 1) 无论病例复杂还是常见, 医生都有可能犯错或失误; 2) 临床医生的知识更新无法与急剧增长的医学知识同步; 3) 对大批量的常规决策工作, 自动化决策效率更高。信息技术的发展以及临床医疗中客观需求的存在, 促进了临床决策支持系统 (Clinical Decision Support System, CDSS) 的发展。

CDSS是将临床信息作为输入, 智能地辅助临床医疗人员形成决策的软件。它一般包括四个基本组成: 推理机、知识库、解释模块及工作存储。推理机利用表述于知识库中的知识及患者的相关资料得出针对性的结

论。由此可见, 临床知识库的构建是保障CDSS运行必须首要解决的问题。

当今, 对知识库系统的开发已经形成了共识, 即知识库系统的开发过程应被视为一个建模活动<sup>[2]</sup>。建立知识库系统意味着建立一个具有专家能力的求解问题的计算机模型。

SAGE (Standards-Based Sharable Active Guideline Environment) 模型是众多医学知识模型中较为有效的一种模型。本文对SAGE模型作了分析, 并以代谢综合征诊疗知识为例, 阐述了建模方法, 针对SAGE表达能力的不足, 提出了改进措施。继而将代谢综合征诊疗的SAGE模型在CDSS中进行应用, 并对CDSS开展实施评估。最后, 提出了SAGE模型仍需改进的方向。

### 1 SAGE模型分析

从20世纪90年代最早的医学知识表达标准Arden Syntax<sup>[3]</sup>发展至今, 国际上已经公开发布了EON<sup>[4]</sup>、PRODIGY<sup>[5]</sup>、GLIF<sup>[6]</sup>等十余种医学知识模型。其中, 2002年发布的SAGE<sup>[7]</sup>医学模型是斯坦福大学医学院联合多个组织的共同研究成果, 该模型设计建立在

收稿日期: 2008-09-26

基金项目: 中国高技术研究发展 (863) 计划资助项目 (2006AA02Z348)

通信作者: 叶枫, 男, 教授, 主要研究方向: 决策支持系统、地理信息系统、管理信息系统。E-mail: yef\_tao@mail.hz.zj.cn

以往模型 (Asbru、GEM、GLIF3、EON、PROformal、GUIDE、PRODIGY) 的基础上。SAGE模型在工作流集成、从异构的临床信息系统中获取患者信息以及模型共享等方面开展了诸多研究, 是较为合适有效的医学知识表达模型。

语法和语义的规范化表达是医学知识模型的设计需求之一<sup>[8]</sup>。SAGE模型引入了SNOMED临床术语、LOINC等术语国际标准。对医学术语采用统一的国际编码, 可以避免理解的不一致, 为实现语义集成打下了基础。

医学知识表达的传统方法是以自然语言为载体, 以非结构化或半结构化的形式出现。这种表达方法, 使知识的协作创建和共享利用存在障碍。SAGE模型将本体论<sup>[9]</sup>应用于医学知识领域。它是结构化的, 并按照本体论的方法, 通过定义一系列的类、槽、实例以及相关的约束来表达医学知识。其中, 实例是类的对象, 槽是对类的属性描述。例如, SAGE模型中的Activity\_Graph类用来表示临床活动的流程图, 描述一段时间内决策和行为的处理流程, 它由description槽 (流程说明)、start槽 (流程起点)、steps (步骤)、transitions槽 (步骤顺序关系) 等构成。该类中, steps槽的值又必须是Activity\_Graph\_Node类 (流程图结点) 的实例。限于篇幅, 图1<sup>[10]</sup>仅列举了Activity\_Graph类中的一部分槽。

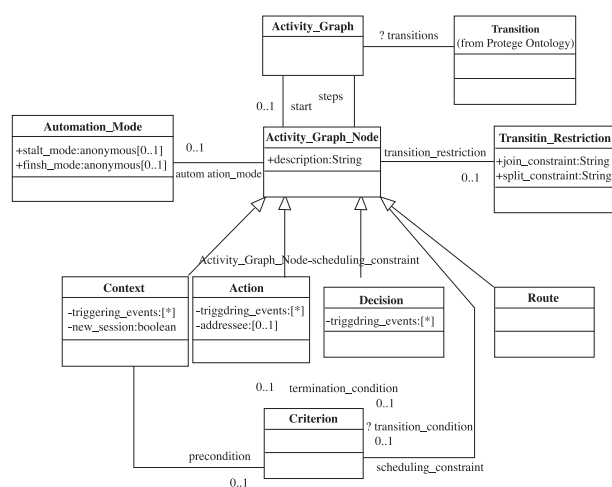


图1 Activity\_Graph类的结构 (部分)

Fig.1 Structure of the Activity\_Graph class

类和类之间还具有层次关系。父类具有抽象性, 为了区别相似类之间的特性, 可对父类进行继承和扩展, 定义子类, 甚至多个层级地定义子类的子类。例如, Action类用来描述行为, 该类派生了: Conclude类 (下达诊断结论) 和External\_Act类 (需要与用户交互

或者与临床信息系统交互的行为) 等, 而Notify类 (通知用户消息)、Inquire类 (向用户询问信息) 等则又是External\_Act类的子类。

理想的医学知识模型, 还应该能够支持单元化建模, 这样可以减少重用、维护模型时额外的负担, 也方便多人协作建模。SAGE模型采用定义类的实例的方法, 可以具体定位到局部的知识表达。当某一部分知识发生更新时, 只需要在局部范围内修改或删减涉及的几个实例, 或新增必要的实例即可, 而无需全局进行修改。而且, 当要表达同一知识内容时, 只需引用同一实例即可, 无需重复定义实例。这样可以减少建模耗时, 也能避免知识更新维护时可能引起的不一致性现象。

SAGE模型的特点还在于, 它能够与临床工作流集成, 采用事件驱动的模式。以往模型在与工作流集成时不成功的原因大多归为三类<sup>[7]</sup>: 1) 在模型外部定义工作流场景; 2) 将集成视为工作流与整个模型的适配; 3) 试图细化工作流的所有元素。SAGE模型与工作流集成采用在医疗过程的合适时机提供决策支持的策略, 而非控制整个工作流, 因此可起到对临床工作流的有效决策支持。

## 2 代谢综合征SAGE建模

代谢综合征是指以胰岛素抵抗为基本病变, 糖调节受损、高血压、血脂紊乱、肥胖和高尿酸血症等多代谢异常在人体聚集的一种状态<sup>[11]</sup>。代谢综合征具有发病率高而知晓率、控制率低的特点, 因此开展早防早治, 对提高全民健康水平具有重大意义。为了说明建模的方法和模型中存在的问题, 以下对医学知识的描述做了删减, 这样与实际创建的模型略有不同。

以代谢综合征表现之一的高血压为例, 在评定高血压患者的危险度时, 需要根据血压分级情况以及具有的危险因素个数综合判断。对于1级高血压患者, 当没有其他危险因素时, 诊断为低危; 当具有1~2个危险因素时, 为中危; 当超过2个危险因素时, 为高危。

采用SAGE建模时, 首先创建一个Activity\_Graph类的实例表示高血压诊断、分级和危险度分层的流程图。然后, 创建一个Decision类的实例表示危险因素判断, 并添加到Activity\_Graph实例的steps槽中。危险因素判断的decision\_model槽值中列入3个Alternative\_Choice类的实例, 表示危险因素个数不同情况下的判断分支。根据SAGE模型定义, 表1列出Alternative\_Choice类的结构, 该类用来描述进行一项决策判断时, 一组匹配条件和否决条件的约束, 与选择采取的行为

之间的对应关系。

表1 Alternative\_Choice类的结构  
Tab.1 Structure of the Alternative\_Choice class

槽名	多重性	类型
alternative	1	Activity_Graph_Node实例、 Decision_Map_Node实例
label	0..1	字符
recommendation _threshold	0..1	整数
rule_in	0..*	Criterion实例
rule_out	0..*	Criterion实例
strict_rule_in	0..*	Criterion实例
strict_rule_out	0..*	Criterion实例

对于第一个判断分支,将alternatve槽值设为Conclude类的一个表示高血压低危的实例,strict\_rule\_out槽值中列入各种其它的危险因素(高龄、吸烟、血脂异常、早发心血管病家族史、缺乏体力活动,……)作为否决条件,表示只要不满足所有这些危险因素,那么给出诊断结论为低危。

对于第三个判断分支,将alternatve槽值设为Conclude类的一个表示高血压高危的实例,strict\_rule\_in槽值中列入各种其它的危险因素作为匹配条件,并且将recommendation\_threshold槽值设为3作为阈值,表示具有这些危险因素的个数超过2个时,那么给出诊断结论高危。

然而,对于第二个判断分支,SAGE模型无法表达“具有1-2个危险因素时诊断为中危”这样的语义。为此,需要对SAGE模型进行扩展,将recommendation\_threshold槽的名称修改为recommendation\_threshold\_min,该槽仍用来表示最小阈值,并且新增加一个recommendation\_threshold\_max槽表示最大阈值。按照扩展后的SAGE模型,将alternative槽值设为Conclude类的一个表示高血压中危的实例,strict\_rule\_in槽值中列入各种其它的危险因素作为匹配条件,并且将recommendation\_threshold\_min槽值设为1, recommendation\_threshold\_max槽值设为2。这样,就能够满足语义表达要求,解决了原SAGE模型表达能力的一个不足。

3 模型应用

根据上一节中的建模方法,可以完成对代谢综合征诊断治疗的知识建模。但该模型不能直接用于推理,无法向医生提供计算机辅助决策建议。可以通过规则生成器<sup>[12]</sup>,将模型自动转换成一组符合CLIPS (C Language Integrated Production System) 格式的推理规

则,形成规则化的知识库。这样,当SAGE模型中定义的临床医疗事件发生时(如患者前来就诊),则CDSS中的推理机可以被驱动。CDSS从医院信息系统中获取患者的个人信息作为推理所需的事实,根据知识库中的规则,产生推理结果,为临床医疗人员提供决策建议。

作为医疗信息融合与临床支持系统课题中的一个研究内容,我们开发了一个CDSS来辅助临床医生进行医疗决策。该CDSS建立了针对代谢综合征诊疗的SAGE模型,以CLIPS作为推理机,以真实的病例作为患者信息获取来源,并在医院中进行实施,对CDSS的临床应用价值进行了评估。评估结果表明,CDSS能够根据患者个体情况差异,在需要时将代谢综合征诊疗知识有选择性地区别呈现给医生,可以建议还需执行某些检验,或诊断是否患有某项疾病,病情的危险度等。

4 结论与展望

本文建立了代谢综合症诊疗的SAGE模型,并进行了适当改进,继而在该改进的SAGE模型的基础上构建了一个实际可用的CDSS。尽管SAGE模型的表达能力存在不足,但可以得到改善。通过临床评估该CDSS,表明SAGE模型有效且通用可行。

但SAGE模型还存在一些问题未能很好解决。例如,临床医生更习惯于自然语言描述,他们无法直接将医学知识进行SAGE建模,而是需要知识工程师与临床医生的配合才能完成;知识工程师需要与临床医生沟通理解医学专业知识,建模表达后的医学知识还需要医生确认一致性,整个建模过程需要耗费较多的时间、人力。又如,SAGE模型对模糊知识的表达能力还不充分,而医学知识不是精确的学科,具有经验性。医生在实际临床医疗决策时,结论也常常具有非确定性。因此,还需要对目前的模型进行深入研究以及完善,以更好地促进医疗质量的提高。

参考文献

[1] Jan H van Bommel, Mark A. Musen主编, 包含飞, 郑学侃主译. 医学信息学. 上海: 上海科学技术出版社, 2002  
[2] 朱承, 曹泽文, 张维明. 知识库系统建模框架的发展与现状[J]. 计算机工程, 2002, 28(8): 3-5  
[3] Geogre Hripesak. Writing Arden Syntax Medical Logic Modules[J]. Computers In Biology and Medicine, 1994, 24(5): 331-363  
[4] Mark A. Musen, Samson W. Tu, Amar K. Das, et al. EON: A Component-Based Approach to Automation of Protocol-directed Therapy[J]. J Am Med Inform Assoc, 1996, 3(6): 367-388  
[5] Peter D. Johnson, Samson Tu, Nick Booth, et al. Using Scenarios[J] in Chronic Disease Management Guidelines for Primary Care. Proc. AMIASymp, 2000, 389-393  
[6] Mor Peleg, Aziz A. Boxwala, Omolola Ogunyemi, et al. GLIF3: The Evolution of a Guideline Representation Format[J]. Proc.



- AMIA Symp, 2000, 645-649
- [7] Samson W. Tu, James R. Campbell, Julie Glasgow, et al. The SAGE Guideline Model: Achievements and Overview[J]. J Am Med Inform Assoc., 2007, 14 (5): 589-598
- [8] Arnoud A.F. van der Maas, Arthur H.M. Ter Hofstede, A. Johannes Ten Hoopen. Requirements for Medical Modeling Languages[J]. J Am Med Inform Assoc., 2001:8:146-162
- [9] Natalya F. Noy, Deborah L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report
- SM1-2001-0880, 2001
- [10] Samson Tu, Julie Glasgow. SAGE Guideline Model Technical Specification, SAGE Consortium, 2006
- [11] 田慧, 代谢综合征的病因和防治原则[J]. 糖尿病天地, 2005, 48(8): 473-475
- [12] Binfei Wu, Xudong Lu, Huilong Duan. An Automatic Knowledge Acquisition Mechanism for Independent Inference Engine Module of CDSSA. Bioinformatics and Biomedical Engineering, 2008, ICBBE 2008 The 2nd International Conference on 16-18 May 2008 Page(s): 1293-1296

## 美国《时代》周刊评出2008年十项医学突破

近日美国《时代》周刊评出了2008年的十大医学突破。

### 1. 利用ALS患者的干细胞生产出运动神经细胞

哈佛大学和哥伦比亚大学的研究人员从两个患有肌萎缩侧索硬化症(ALS)的老年妇女身上提取出干细胞,生产出了第一个运动神经细胞。通过使用该细胞,可在培养皿中观察细胞的发育、衰退和死亡,以研究该疾病的进展情况。

### 2. 炎症是心脏病发作的“触发器”

美国波士顿布莱根妇女医院的鲍尔·瑞德科确认了心脏病中一个独立且功能强大的风险因素,那就是炎症。它起着关键的“触发器”的作用,可增加脂肪斑的不稳定性,使其更有可能破裂、阻塞血管,最后导致心脏病发作。

### 3. 无疤痕手术

通过进入口腔、阴道和结肠等“自然腔道”直接完成手术,无需在皮肤上进行切口。2008年3月,美国加州大学圣迭戈分校的科研小组通过阴道成功为一名妇女进行了阑尾切除术。该技术对一些胃旁路患者也大有裨益。该手术是通过微小的切口穿过人体内部组织,这样可降低病患的疼痛感和感染风险,使患者在最短时间内恢复健康。

### 4. 基因组图谱绘制走近大众

现在,只需一些唾液,就能绘制出个人基因组的完整图谱,并揭露其中隐藏的奥秘。科学家从唾液提取DNA,对其复制并在90%的已知遗传变异中进行查找。这些变异则代表了从乳糖不耐受症到前列腺癌等不同的特征或情况。科学家有望通过这些遗传信息,预知人类可患疾病,以便及时进行预防治疗。

### 5. 引发早老年痴呆症的新基因被发现

近期发现了4种可引发早老年痴呆症的新基因,有望在临床解剖之前,对此种病症进行诊断。这些基因或许与神经元的死亡相关。随着疾病的不断恶化,脂质斑块和蛋白质纤维缠结将渐渐堆积在大脑中,不断对神经元进行破坏,并最终导致其死亡。新基因或可为科

学家了解神经元如何保持活性提供帮助,并为未来的早老年痴呆症治疗奠定基础。

### 6. 五合一儿童复方疫苗上市

美国食品及药物管理局批准了首个五合一儿童复方疫苗潘塔塞尔(Pentacel)的上市。该疫苗可用于婴幼儿预防白喉、破伤风、百日咳、脊髓灰质炎和B型嗜血性流感杆菌感染。目前已在5000多名婴儿身上得以应用,接种的儿童仅出现包括发烧、注射处红肿等轻微副作用。

### 7. 基因扫描探测乳腺癌

基因扫描检测乳腺癌,可查明患者的哪些基因能最有效地对癌症治疗药物赫赛汀(Herceptin)敏感。用于治疗可释放大量HER2的转移性乳腺癌。此外,利用一种名为OncotypeDx的基因测试方法,还可确定乳腺癌复发的风险,并通过化疗药剂最有效地杀死一种特殊的肿瘤。

### 8. 血液测试可确诊唐氏综合征

通过提取准妈妈的血液样本来确诊婴儿是否患有唐氏综合征。这种测试方法可检测出母体血液中第21对染色体出现的三体现象,从而确诊婴儿是否患有相关疾病。

### 9. 治疗化疗副作用的药用贴片问世

2008年9月,美国食品及药物管理局批准全球第一种经皮吸收的5-HT-3受体阻断剂Sancuso(Granisetron,格拉司琼),用于治疗化疗引起的恶心呕吐。这种药用贴片可附着在皮肤表面连续释放一定量的止吐药物格拉司琼,从而减轻患者的恶心、呕吐等症状。一经敷用,可连续5天抑制反胃和呕吐现象。

### 10. 成人干细胞培育器官移植首获成功

2008年6月,英国、意大利和西班牙三国组成的研究小组首次利用由部分成人干细胞培育而成的气管为一名30岁的西班牙女患者实施了器官移植手术。手术中未使用免疫抑制剂等药物,术后4个月病人未出现排斥现象。这次手术的成功标志着干细胞技术的重大突破,具有重要的里程碑意义。

(本刊讯)