

基于临床数据中心的医院知识服务系统构建

景慎旗

刘云 缪姝妹 张小亮

(1 南京大学信息管理学院 南京 210023

(江苏省人民医院 南京 210029)

2 江苏省人民医院 南京 210029)

〔摘要〕 基于当前信息服务领域呈现的信息服务向知识服务转变的新态势, 提出构建基于医院临床数据中心的知识服务系统。介绍系统的体系架构, 分析系统特点, 阐述系统应用效果并对未来知识服务系统进一步发展需要关注的问题进行讨论。

〔关键词〕 价值链模型理论; 临床数据中心; 知识服务系统

〔中图分类号〕 R-056 **〔文献标识码〕** A **〔DOI〕** 10.3969/j.issn.1673-6036.2017.02.018

Construction of Hospital Knowledge Service System Based on Clinical Data Center JING Shen - qi, 1 School of Information Management, Nanjing University, Nanjing 210023, 2 Jiangsu Province Hospital, Nanjing 210029, China; LIU Yun, MIAO Shu - mei, ZHANG Xiao - liang, Jiangsu Province Hospital, Nanjing 210029, China

〔Abstract〕 Based on the new trend of transition from information service to knowledge service currently emerging in the field of information service, the paper presents the construction of a knowledge service system based on clinical data center. It introduces the architecture of the system, analyzes its features, describes its application effect and discusses the issues to be concerned for the further development of the knowledge service system in the future.

〔Keywords〕 Value chain model theory; Clinical Data Repository (CDR); Knowledge service system

1 引言

当前在信息技术迅速发展的推动下, 传统的信息服务机构和网络信息服务运营商逐步走上由信息服务向知识服务的转型之路^[1-3]。知识服务是面向用户的服务, 是基于用户文献信息的需求而开展的

服务, 相比传统的信息服务, 能够提供价值密度更大、更能揭示事物本质、更加精准的认知集合, 高效、有力地辅助人们在社会生活中积极开展生产实践, 提高生产实践的效率与质量。临床数据中心 (Clinical Data Repository, CDR) 是目前医院信息化建设的重点, 也代表着今后医疗信息化应用的一个方面, 其有力整合了医院以患者为中心的多维度全生命周期的诊疗数据, 是医院各业务系统的数据集合, 具有很高的应用价值^[4-5], 深度挖掘意义重大。近年来, 信息价值链模型理论被广泛用于数据挖掘中, 其将客观事实通过数据、信息中间体的多

〔修回日期〕 2016-10-31

〔作者简介〕 景慎旗, 博士研究生, 发表论文 4 篇; 通讯作者: 刘云, 博士, 主任医师, 处长。

次转换, 经历提炼萃取、抽象概括, 最终由大量事实描述转化为高度精炼和概括的知识并以知识服务的形式对外供用户消费使用^[6]。鉴于此, 本文参照信息价值链模型理论构建基于临床数据中心的医院知识服务系统, 通过对医院海量的数据进行深度挖掘, 总结提炼并发现知识, 以为医院各类医务人员提供多样性、启发性的知识服务, 实现智能辅助决策和临床科研智能支持等多种智能应用。

2 体系架构 (图 1)

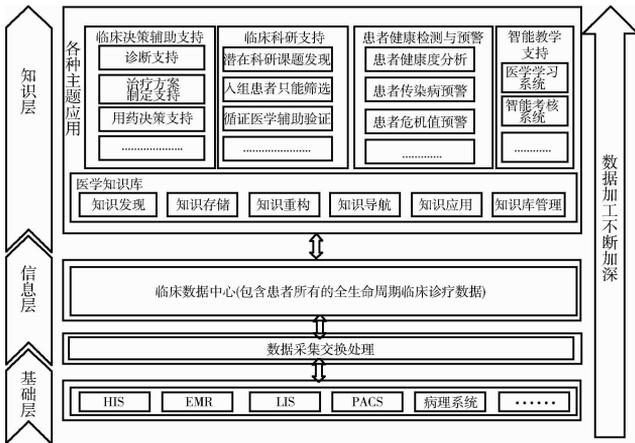


图 1 基于临床数据中心的医院知识服务系统架构

2.1 基础层

基础层为信息价值链模型理论中的数据层, 处于整个价值链的上游。该层主要集中了医院各业务系统的数据, 是保障医院各业务正常运行的基石, 同时也是整个知识服务系统的原始数据来源, 为知识服务系统挖掘发现知识提供了广泛、丰富的原材料。因此该层的数据多样性、丰富度、完整性、准确性以及一致性等决定了知识层是否能够有效开展知识发现活动。数据质量越好, 最终越能够发现新知识, 知识创新的可能性越大。

2.2 信息层

信息层处于整个信息价值链的中游, 整合处理上游分散在各系统的临床业务数据, 进行初加工后统一存储, 为下游知识层实现医学知识发现提供海量有效信息, 其在知识服务系统中起着承上启下的

重要作用。信息层主要包括临床数据中心以及数据采集交互平台, 其中临床数据中心是数据的统一存储地, 其数据直接供知识层进行深度挖掘, 是实现知识发现的直接基础。该层的工作重点是: 对采集进入临床数据中心的数据实施有效管控: 进行预处理, 包括数据清洗、转换、标准化等工作, 确保采集到高质量的数据, 尽可能为知识层实现知识发现提供规范、纯净的数据。其数据存储格式包括结构化、半结构化和非结构化。

2.3 知识层

2.3.1 概述 知识层处于整个信息价值链的下游, 是整个知识服务系统中最复杂、最核心也是应用价值最高的一层, 深层次的知识发现和知识创新均在该层完成; 同时该层直接面向用户, 向用户提供经萃取精炼、挖掘重组的各类医学知识, 是知识服务的最前端展示, 与用户关系最为紧密。该层中最核心的部分是医学知识库, 该知识库是基于临床数据中心提供的海量数据构建的, 是知识服务系统的主要载体, 其主要有如下 6 大功能模块。

2.3.2 知识发现模块 该模块在临床数据中心的海量数据中深入挖掘发现知识, 是初级信息服务向高级知识服务转变的关键环节: 一方面对大量纷繁复杂的数据进行抽象提炼升华, 剥离个性的内容, 将其中共性的内容高度概括, 升华为知识; 另一方面通过对海量数据进行深入关联分析, 发现隐含其中的内在的、潜在的关联关系, 总结提炼为知识。其具体实现方式有多种, 既可以采用传统的数据挖掘方法, 如分类法、聚类法、关联分析等进行知识深度挖掘, 也可以采用流行的神经网络算法等深度学习算法进行自主学习发现知识^[7]。该功能模块强调主动感知与智能调节的“智慧”性: 因为大部分知识发现都是无预设条件的, 是通过不断迭代学习感知挖掘方向与重点, 不断自我调节和修正, 从而完成最终的知识发现目标。同时利用算法尽可能将发现的知识进行降维和细化, 即将知识的颗粒降到最低, 成为不可再分的知识点——所谓的知识元。细颗粒度的知识有利于后期进行知识的重组再加工。

2.3.3 知识存储模块 该功能模块主要将发现的

知识进行存储和管理。针对医学知识库的海量知识,可采用分布式存储的方式完成。面对知识库增加新知识、更新陈旧知识以及摒弃过时或错误知识的特点,从管理制度上和技术上构建知识库变更规范,确保知识库的任何知识变更能够得到可靠有效的执行。

2.3.4 知识重构模块 利用该模块可以实现知识创新:将知识发现产生的知识元进行知识重构,生成各层次、各类别的医学知识,满足不同对象、不同应用场景的需要,同时在进行知识元重构时有可能发现新的知识、意想不到的知识,这是知识重构模块最吸引人之处。该模块在知识创新方面有着极其重要的应用价值。

2.3.5 知识导航模块 该模块通过生成医学知识图谱以及各医学知识点之间的关联路径实现知识学习和检索的可视化导航——知识应用的可视化。将各类医学临床知识内部、潜在的关联关系显性化表示能够显著降低知识学习的难度,满足医务人员、医学院学生甚至患者等不同等级、不同知识背景的学习对象不同程度的学习需求。用户可通过对知识点可视化操作学习该知识点以及与其相关的周边知识点,提高知识服务的精准度与便捷度。

2.3.6 知识应用模块 大型医院一般集医教研应用于一体,具有复杂多样的需求,知识应用模块能够很好地满足这些需求:通过可视化操作或者计算机编程等方式将知识点根据应用场景进行整合产生各类主题应用以满足不同场景的需求,如临床决策支持、临床科研支持、患者健康监测与预警、医学教学等^[8-9]。相比较知识导航模块提供的初级、简单的知识服务,知识应用模块所提供的知识服务更为高级、复杂和多样:知识应用模块通过建立基于主题的关联关系将底层知识点进行整合,构建复杂的基于主题背景的知识网络,从而实现宏观场景展现和主题应用。这种方式无须像传统的数据集市一样将既有知识按照主题要求重复建设、存储,显著降低了系统数据的冗余性和数据容量,提高数据利用率和可复用性,应用更加便捷、灵活。

2.3.7 知识库管理模块 该功能模块为系统管理员提供系统的日常维护功能以及常规问题解决方

案,帮助系统管理员管理系统,确保系统安全、可靠、高效的运行。

3 系统特点

3.1 知识创新性和启发性

本知识服务系统的知识重构模块利用细颗粒度的医学知识元在预设任何前置条件下进行各种方式的重构、关联与整合,产生的结果具有很强的不确定性和多样性,利用海量知识元可以重构产生海量结果,其中很多是前所未有的。通过对这些结果进行筛查检验,可以从中发现少量但极具启发意义、能够驱动用户展开更加深入的临床研究的创新知识点,这对于强调医学研究的原创性与新颖性是非常有意义的。

3.2 知识深度融合性

由于本知识服务系统的数据源覆盖全院各类系统,通过临床数据中心进行业务数据的整合——实现数据在空间维度和时间维度的高度融合,可为知识层实现高质量的知识发现提供坚实的数据基础:产生的知识具有全面性、多维度、多层次、深度融合等特点,避免因片面、孤立的数据源而产生的错误、曲解的知识。本知识服务系统提供的深度融合的医学知识能够满足各类需求,使用户更加全面、客观、正确地理解认识不断涌现的新医学现象。

3.3 实现医学知识的“显性化”和“组织化”

很多专家个性化的临床知识带有较强的个人色彩,遍布在平时的诊疗活动中,具有片断性、离散性和偶然性特点,难以用语言准确地描述。而这些知识都是宝贵的医学财富,对于提高医学理论和实践水平都有积极作用,因此具有很强的传承意义,必须进行显性化、组织化管理,提升价值^[10]。通过本知识服务系统中的知识发现和知识重构模块挖掘发现这些隐藏其中的“隐性”、“个性”的医学知识,可将其准确阐述并在统一组织下妥善保存,避免因个人因素导致这些医学知识的消失,不断发扬光大,提升价值。

3.4 促进知识的流通与共享, 满足个性化的医学知识需求

本知识服务系统可以将某个患者的局部医学知识信息提取后泛化处理变成普适性的医学知识提供给其他用户。通过这种方式实现医学知识资源的有效供给, 解决医学知识资源分布不均的局面, 提高知识复用性, 实现知识增值, 体现知识共享性。同时针对不同层次、不同知识背景的用户, 鉴于需求的个性化和差异性按需提供相应的知识服务, 满足用户个性化需求, 提高知识服务与个体吸收的契合度。

4 应用效果

4.1 为临床诊疗提供辅助决策

通过对海量离散化、碎片化的诊疗数据进行深度挖掘, 重构后可固化、显化蕴藏其中大量有意义的诊断与治疗知识, 这些知识来源于实际诊疗活动而非教科书中, 具有显著的应用价值。本系统中的临床决策辅助支持应用将这些知识点显式提供给医务人员, 可为其诊疗方案的制定提供切实有效的临床决策建议, 有效保障医务人员, 特别是低年资医务人员的诊疗决策质量, 同时也藉此提升其业务水平。

4.2 为临床科研提供新思路

临床科研支持应用通过对既有的海量临床诊疗资料进行分析挖掘, 可发现其中潜在的、有科研意义的临床科研兴趣点, 这些兴趣点由于隐藏较深, 一般难以用传统的方法发现。这些兴趣点为临床科研提供突破口, 同时通过系统快速筛选功能为科研提供有效的、高质量的人组患者, 显著提升临床科研的效率与质量。

4.3 提高医学教学质量与效率

本系统中的智能教学支持应用可为各类医学背景的学生、进修生等人员提供医学自助学习服务。由于该应用中相关的医学教学案例来源于真实的患者诊疗事件, 因此具有较强的真实性与临床可操作性, 更加贴近现实临床应用。通过将医学知识点

与相关真实案例相结合, 从理论与实践两个角度向学习人员系统阐述知识点, 有效提升学习人员对知识点理解的透彻度, 进而提升医学教学质量和效率。

5 讨论

5.1 提高“人机交互”的深度和知识服务和谐度

一方面知识服务系统不同于传统的应用系统, 具有较强的自主学习能力, 同时也需要一定程度的人工干预以确保自主学习效果最大化, 因此对于系统管理者要求高, 要求其有充分的知识储备, 包括医学知识、计算机知识、数学知识等, 才能够很好地管理该系统, 对系统学习性能进行优化, 提高系统的知识服务水平。另一方面要求系统用户具备一定的信息素养和业务知识: 系统与用户构成整个体系, 二者缺一不可, 任何一方成为短板都会影响最终的应用效果。用户如能很好地应用该系统, 则能从系统提供的知识服务中快速、准确地获取自己需要的知识, 进而进行知识创新。

5.2 重视患者隐私保护, 强调信息安全

临床数据中心存储的数据全面描述了患者全生命周期的健康和患病情况, 具有高度的隐私特征。为保护患者隐私权不被侵犯, 必须考虑在进行深度数据挖掘和知识发现活动中遵循相关的法律法规, 采用一定技术方法确保患者各类隐私信息不被泄露, 切实保护患者隐私权。美国的患者隐私保护工作起步早, 成果较大, 《健康保险携带和责任法案》(Health Insurance Portability and Accountability Act, HIPPA) 是其医疗领域有关患者隐私保护的一项重要法案, 对患者隐私保护起到了重要作用。我国在这一领域起步较晚, 相关立法有待完善, 因此 HIPPA 对于我国制定相关患者隐私保护法案具有重要参考意义。而各类匿名化技术如泛化技术、替换技术和加密技术在不影响统计学意义上的分析结果的同时对患者隐私信息进行脱敏, 有效保护患者各种隐私信息不被泄漏。

5.3 进一步实现知识发现方面的创新

知识服务创新与发展体现在理论与实践两方面

的突破。目前知识发现基本由系统自动完成,主要是基于机器学习(Machine learning)实现的。在机器学习领域存在两条理论路线:一条是传统的工程学算法理论路线;一条是模拟人脑思维的认知算法理论路线。前者取得了很大的成绩,但由于与人脑思维有着本质的区别,无法实现真正意义上的类人脑思维模式,而后者的理论起源于人脑思维理论,目前尚未取得突破性进展。目前学术界也非常关注这一重大问题,相关研究热点也非常多,一旦这方面有革命性突破,将会对知识发现的效率和准确性产生深远影响,或许能从根本上提升知识服务系统的性能与知识服务提供水平。

5.4 鼓励知识共享,实现知识增值

在当今知识经济和“互联网+”时代,日益强调“共享”精神,鼓励知识的扩散与传播,降低知识传播与获取的门槛。相比石油、粮食等传统的有形资源存在传播或使用损耗的问题,知识最大的特点在于其在扩散、传播和应用的过程不但不会损耗消失,反而会不断增值——通过多人多次对知识的不断检验修正,知识的内涵和应用价值不断提升。然而仅依靠知识服务系统来实现医学知识共享、复用尚显单薄,因此需采用制度管理与系统应用相结合的方式,实现知识共享常态化、常规化,为医务人员、医学院学生、患者等不同用户群体提供低门槛的医学知识传播与应用途径,实现医学知识共享,促进医学知识利用的最大化。

6 结语

临床数据中心作为医院各业务数据的总集成存储,全面真实地保留了临床诊疗活动的各项纪录,因此具有极高的深度数据挖掘分析及应用价值。本研究利用临床数据中心构建医院知识服务系统,通过对原始大数据进行深度分析、深加工后提炼出高

度浓缩、具有普遍意义的医学知识点集合,为医院的医生、进修生等医学领域专业工作者提供可视化的知识服务。本知识服务系统具有知识创新与深度融合的特点,将医学知识“显性化”和“组织化”,针对个体需求的差异性提供个性化知识服务并取得较好的效果,这是在知识服务领域一次有益的探索与尝试。在今后的工作中需要不断完善系统,提高知识服务的便捷性与敏捷性,满足用户不断增长的多样性及个性化需求。

参考文献

- 1 代涛,钱庆,王小万,等. 医疗卫生领域知识服务与知识管理的理论和实践[J]. 医学信息学杂志, 2008, 29(4): 1-10.
- 2 黎艳. 信息服务向知识服务转变的探析[J]. 图书情报工作, 2003, (2): 31-34, 25.
- 3 马利,崔志伟,毛树松,等. 我国医学知识库应用现状研究[J]. 医学信息学杂志, 2013, 34(11): 55-59.
- 4 姜小娟,刘歆农,张民,等. 基于 CDA 标准的临床数据中心研究和应用[J]. 中国数字医学, 2014, 9(5): 99-101.
- 5 徐晓辉. 医院临床数据中心构建探讨[J]. 中国数字医学, 2015, (2): 22-24.
- 6 李后卿,董富国,郭瑞芝,等. 信息链视角下的医学信息学研究的重点及其未来发展方向[J]. 中华医学图书情报杂志, 2015, (1): 1-5.
- 7 李伟,赵卫利,刘冠群,等. 知识库和知识发现技术在知识服务方面的应用[J]. 现代情报, 2009, 29(7): 180-182.
- 8 Sim I, Gonn P, Greenes RA, et al. Clinical Decision Support Systems for the Practice of Evidence-based Medicine[J]. Journal of the American Medical Informatics Association, 2001, 8(6): 527-534.
- 9 傅静萍. 知识服务在临床医疗中的应用[J]. 科技情报开发与经济, 2010, 20(34): 121-123.
- 10 律苗,龙莉艳,张桂云,等. 医学科研团队的知识服务模式探讨[J]. 中国医院, 2012, 16(5): 64-66.