

大数据背景下医学数据挖掘的研究进展及应用

秦文哲¹ 综述 陈进¹ 董力² 审校

1. 四川大学华西医院 循证医学与临床流行病学教研室; 2. 心血管外科(成都 610041)

【摘要】 大数据时代的来临,带来了生活、工作与思维的大变革。在医学领域,21世纪以来,随着社会经济的发展 and 医疗卫生的进步,人类疾病谱正在发生改变,疾病种类不断增多,病因、诊断及治疗的复杂性也在逐渐增加。为提高人类的健康水平,探索疾病的发生、发展规律,需要我们不断探究,通过对医学大数据的充分挖掘和利用,从海量的知识中发现隐藏其中的规律。这对于提高医学信息管理水平,为诊断和治疗疾病提供理论和方法学的支持,促进临床实践和决策将发挥重要的作用。本文以大数据为背景,对大数据的由来、意义、分类、特征、数据挖掘和医学数据挖掘的研究进展、应用和未来发展进行综述。

【关键词】 大数据;数据挖掘;医学研究

Progress and Application of Medical Data Mining under the Background of Big Data

QIN Wen-zhe¹, CHEN Jin¹, DONG Li²

1. Department of Evidence-Based Medicine and Clinical Epidemiology, West China Hospital, Sichuan University, Chengdu 610041, P.R. China;

2. Department of Cardiovascular Surgery, West China Hospital, Sichuan University, Chengdu 610041, P.R. China

Corresponding author: CHEN Jin, Email: ebm_chenjin@126.com; DONG Li, Email: donglikn199@163.com

【Abstract】 The era of big data has brought a big revolution that will transform the way we live, work, and think. In medical field, as the development of social economics and medicine since 21 century, the human disease spectrum has been changing, the disease type has been increasing, and the complexity of the etiology, diagnosis and treatment of disease have been gradually increasing. In order to improve the healthy level, and explore the law of disease occurrence and development, we should constantly research to find discipline in enormous knowledge by fully mining and using the big medical data. It will be helpful to improve the level medical information management. And it can be supportive to the diagnosis, treatment, clinical practice and decision-making. We did the review under the background of big data, and the main content of this review is about the origin, meaning, classification, features of big data as well as the research process, application and future development of data mining, especially clinical data mining.

【Key words】 Big data; Data mining; Medical research

【Foundation item】 National Science & Technology Support Program during the Twelfth Five-year Plan Period (2011BAI11B18)

进入21世纪,随着计算机、互联网技术、云计算、移动终端、数据储存方式的高速发展和变化,一个新的时代——大数据时代,正在强烈冲击着我们社会、生产和生活的方方面面,带来了生活、工作与思维的大变革。医学和其他领域一样,在经历了仅依赖经验、理论、假设和价值观去发现未知世界的规律的“无数据时代”和通过有意识地收集数据,依赖抽样数据、局部数据和片面数据,支持相关的判断和决

策的“样本数据时代”后,开启了一次重大的时代转型——大数据时代。如何充分认识、挖掘和利用医学大数据,从中发现有价值的信息,使其为临床实践和决策服务,是当前值得深入探讨的课题。

1 大数据时代

1.1 大数据的由来

2002年8月apache.org的开源项目Nutch最早开始引用“大数据”这个术语。当时,大数据用来描述为更新网络搜索索引需要同时进行批量处理或分析的大量数据集。网上交易、电子邮件、视频文件、搜索查询、健康档案、社交互动、科学数据以及传感

DOI: 10.7507/1007-4848.20160013

基金项目:“十二五”国家科技支撑计划项目(2011BAI11B18)

通讯作者:陈进,Email: ebm_chenjin@126.com; 董力,Email: donglikn-199@163.com

器和移动电话的应用等都促进了大数据的产生。

人、机、物三元世界的高度融合引发了数据规模的爆炸式增长和数据模式的高度复杂化,世界从此进入大数据(Big Data)时代,吸引了越来越多的关注^[1]。在过去的20年里,数据开始在各领域大规模增加。国际数据公司(IDC)的一份报告显示,在2011年,整个世界上创建和复制数据量是1.8 zb,这相当于全球每个人每天都去做2.15亿次高分辨率的核磁共振检查所产生的数据总量,这些数据在五年内增加了近9倍^[2]。事实上,全球知名咨询公司麦肯锡最早提出了“大数据”时代已经到来。该机构在研究报告中指出,数据已经渗透到每一个行业和业务职能领域,逐渐成为重要的生产因素。

1.2 大数据的定义

大数据是一个抽象的概念,尽管其重要性已经被公认,但各领域对它的定义各不相同,目前比较认可的有如下几种:维基百科认为大数据是指难以用现有的数据库管理工具处理的兼具海量特征和复杂性特征的数据集成。一般意义上,大数据是指无法在可容忍的时间内用传统IT技术和软硬件工具对其进行感知、获取、管理、处理和服务的数据集合^[3]。Apache Hadoop(2010年)指出:在通用计算机可接受的范围内,不能被捕获、管理和处理的数据集。在此定义的基础上2011年5月,麦肯锡集团宣布大数据是下一个创新力、竞争力和生产力的前沿。IDC报告(2011年)^[2]指出:“大数据技术描述了新一代的技术和架构,通过启用高速捕捉,发现和分析技术,从大量的各种各样的数据中提取有价值的信息。NIST(美国国家标准与技术研究院,2012年)认为,“大数据指的是数据的数量、采集速度或者那些无法用传统的方法来进行有效分析的数据,或者是可以用重要的横向放大技术进行有效处理的数据”,它关注的是大数据的技术方面。此外,产业界和学术界对大数据的定义也有相当多的其他讨论。

总之,无论哪种定义,都说明大数据不是一种新产品或新技术,它只是数字化时代出现的一种现象或特征。我们最应该关注的不是其定义,而是它所带来的价值。相比现有的其他技术,大数据具有“廉价、迅速、优化”的优点。通过对海量数据进行分析 and 存储,增加了供人类分析和使用的数据量;同时,数据的交换、整合和分析,可以帮助人类发现新知识,创造新价值,带来“大知识”和“大发展”。

1.3 大数据的分类

依据数字集合的结构状态分类,大数据可分为:(1)结构化数据,如企业的财务系统、人事系统(2)半

结构化数据,如电子邮件、HTML网页(3)非结构化数据,如视频、移动终端、传感器等产生的数据。依据数据的应用类型,也可以将大数据分为海量交易数据、海量交互数据以及将交易和交互数据糅合起来进行处理的海量处理数据三类。依据数据的来源,大数据可分为行政记录数据、商业记录数据、互联网及搜索引擎数据三大类^[4]。

1.4 大数据的特征

大数据的特征一般概括为四个“V”(1)volume即“数量体积巨大”,从TB级跃升到PB级(1PB相当于50%的全美国图书馆藏书量)随着数据的大量生成和收集,数据的规模变得越来越大,已经超出了传统的存储和分析的技术^[5]。(2)velocity即“速度”,也就是大数据的及时性,这意味着数据的收集和分析,必须迅速和及时的进行^[6]。一般要求在秒级时间范围内给出分析结果,以最大化地利用它的价值。(3)variability即“数据类型繁多”,包括半结构化和非结构化数据,如音频、视频、网页和文本,以及传统的结构化数据^[6]。(4)value即“价值”,集中体现在价值密度低,而商用价值高。大数据多为非结构化和半结构化的数据,用于分析时会花费过多时间和金钱,以监控录像为例,连续不间断监控过程中,有用的数据可能仅仅有一两秒^[7]。近年,在四“V”特征的基础上,又增加了数据获取与发送的方式便捷(veracity)、准确(veracity)和繁复(complexity)三个特征。在利用大数据时关键的问题就是如何从一个规模巨大、生成迅速和种类繁多的数据集中发现价值。如果不利用收集到的数据,你只能拥有“一堆数据”而不是“大数据”。综合大数据的定义、类型、价值、特征等维度进行考察可以看出,大数据具有预警性、预测性、差异性、共享性和动态性等特性。

2 数据挖掘

伴随着大数据时代的来临,DM(数据挖掘)应运而生。数据挖掘是指通过对大量有噪声的模糊数据,以及随机的实际应用数据的自动和半自动化分析和探索,来发现其中有意义的模式和规则。最早于1989年举行的第11届AAAI(美国人工智能)协会学术会议上提出,是近年来随着人工智能和数据库技术的发展而出现的一门新兴技术^[8]。

与传统的统计分析技术相比,数据挖掘有着自身的本质特征。DM是在没有明确假设的前提下去挖掘信息并发现知识。DM所得到的信息应具有先前未知、有效以及可实用三个特征^[9]。数据挖掘的出现不是为了替代传统的统计分析技术;相反,它是

统计分析方法学的延伸和扩展^[10]。

DM的应用通常有两大类:预测和描述。预测任务主要是根据其他属性的值,预测特定属性的值,主要有分类和回归2种模式。描述任务的目标是概括数据中潜在联系的模式(相关、趋势、聚类、轨迹和异常),主要有关联分析、聚类分析、异常检测3种模式^[11]。在DM算法的理论基础上,一些重要的数据挖掘技术包括:关联规则法、聚类分析、链分析、决策树、人工神经网络、遗传算法、概率论、数理统计、粗糙集和模糊处理技术等^[12]。

3 数据挖掘在医学中的应用

3.1 数据挖掘在医学实践中的意义

随着信息时代的到来,数据挖掘被越来越多地应用于临床实践。利用信息技术,医疗记录和随访数据可以更有效地被存储和提取。同时,从医学数据中寻找潜在的关系或规律,从而获得有效的对病人进行诊断、治疗的知识;增加对疾病的预测准确性,在早期发现疾病,提高治愈率。数据挖掘对医学实践的重要性主要体现在四个方面:(1)医学数据挖掘会大大增加产生新知识的速度,利用计算机技术分析电子病历中包含的非结构化数据能够更好的进行自动化的数据采集。(2)医学数据挖掘可以帮助进行知识传播,大量的研究使得知识的转化非常困难,大部分临床医生难以跟上最新的证据来指导临床实践。这个问题可以通过分析现有的电子病历产生一个仪表盘来指导临床决策。如IBM的沃森超级计算机与斯隆凯特林癌症中心合作,利用这个方法辅助临床医生对癌症患者做出诊断和提出治疗方案。(3)通过整合系统生物学与电子病历数据,医学数据挖掘可以为个性化医疗计划转变为临床实践提供机会^[13]。(4)通过直接向病人提供信息来转化卫生保健知识,让病人在治疗过程中发挥更积极的作用。

3.2 医学数据挖掘的特点

临床医学地位特殊,医疗结果生死攸关。这使得临床数据库在进行数据挖掘时,有其自身的特点。原始的临床数据数量巨大并且具有异质性,这些数据大多来自于电子病历、医学影像、病历参数、化验结果以及临床医生的观察和解释^[14]。这些临床信息具有多样性、隐私性、冗余性、不完整性、缺乏数学性质等自身的特殊性和复杂性,并且涉及到伦理和法律问题,使得医学DM与常规DM之间存在较大差异。首先,就临床研究而言,基于大数据的临床研究不同于以往的RCT(随机对照试验)。毫无疑问,随机对照试验处于证据金字塔的顶部,可以提供高

水平的证据来证明干预的有效性^[15]。然而,随机对照试验并不能解决所有的临床问题^[16]。随机对照试验本身存在固有的局限性,RCT往往是在理想状态下进行,和真实的临床实际存在许多差异,不能代表日常医疗中遇到患者的真实情况,通常采用抽样的方法获取样本,并且存在严格的纳入排除标准,研究中所包含的病患往往只是一部分符合特定要求的群体,存在样本代表性和外部真实性等问题。大数据的整体性和海量信息特性,弥补了随机对照试验的上述不足,为解决临床问题提供了“真实”的临床环境,基于此“真实世界”所获得的结果和证据更接近临床实际,也更有助于正确的临床决策。其次,由于个体差异的存在,使得医学数据挖掘具有区别常规数据挖掘的本质特征。数据挖掘来自于统计学的概率抽样、估计和假设检验。理论上基本情况相同时,病人对疾病或药物的反应是相似的,而病人个体因为先天或后天因素有少数人会出现在性质和数量上有显著差异的反应,如高敏性、低敏性以及特异质性等反应。因此,医学数据挖掘可以针对病人个体进行个体化预测,给出个体化的治疗方案。其次,医学数据挖掘的另一特点是底层数据结构很难进行数学表示。相较于可以将数据直接代入公式、模型来反应数据之间关系的物理科学,医学数据挖掘没有一个正式的可比的结构供数据挖掘者组织信息。虽然,随着计算机技术的发展以及新的数据挖掘技术的出现这一难题会逐渐被克服,但这仍旧是医学数据挖掘的一个重要特征。再次,医学数据挖掘存在伦理性、法律性和社会性问题。数据的所有权,数据的隐私和安全,预期收益以及行政事务等都使得医学数据挖掘与常规数据挖掘存在不同。

人类疾病的复杂性、健康的重要性和医学数据的特殊性,个人、团体和人群水平上疾病后果的严重性,以及我们处理这些复杂情况的能力促进了数据挖掘技术的进步。同时,数据挖掘技术的进步也可以为一些临床研究提供假设^[17]。然而基于真实世界的临床环境,许多内部和外部的因素,如病人的个体化特征、医疗设施、获得卫生保健的途径等,使得在对临床数据进行分析时,数据挖掘技术的应用相对有限。近年来,ANN(人工神经网络)依靠其高度的并行性、良好的容错性与联想记忆功能以及十分强的自适应、自学习能力,开始越来越多的应用到临床数据的分析中,其主要包括分类、诊断、成像、波形分析和结果预测等。

3.3 大数据背景下的医学数据挖掘应用

3.3.1 预警性 大数据的预警性在于相对于提前设

置的标准,当数据发生异常时,通过一定的机制可以发出警告,从而迅速采取相应措施,及时解决问题。滕琪等^[18]通过运用新兴的尖端云计算技术,设计研发了健康云平台。采用分布式云存储技术存储大规模的异构多模态生理信号数据,将数据挖掘模型(L1-Logistic)整合至 MapReduce 框架中,快速挖掘用户的健康信息以及重大疾病高风险因子,让用户能实时地了解自己的身体状况,同时对用户的异常状况给出预警信息,并通知其前往医院就医,实现了对重大突发疾病的早期预警。麻省理工学院、密歇根大学和一家妇女医院创建了一个计算机模型,用于分析心脏病患者的心电图数据,预测未来一年内患者心脏病的发病几率。通过机器学习和数据挖掘,该模型可以通过累积的数据进行分析,发现高风险指标,从而改变了过去医生由于缺乏对之前数据的比较分析,导致对 70% 的心脏病患者是否再度发病缺乏预判的现象^[19]。

3.3.2 预测性 正如 Viktor Mayer-Schönberger 所说:“预测,大数据的核心”。数据挖掘在临床实践中也更多地被用于预测建模,使用患者的特定信息来预测疾病的结果,辅助疾病诊断和推荐治疗措施,从而支持临床决策。

预测建模(predictive modeling)主要是利用自变量函数的方式为目标变量建立模型,包括 2 种模式:分类和回归。分类是对离散的数据进行预测。在临床医学中,疾病的诊断就是典型的分类过程。刘娟^[20]对数据挖掘的三种分类预测算法(C5.0、BP-人工神经网络和 TAN 贝叶斯网络)进行研究和探讨,构建较为适合的模型,用于胃癌的早期预警、诊断和分类的自动化。回归主要是对连续和有序的数据进行预测,可以广泛的应用于疾病的诊断、预后的判别和药物剂量预测等。如 Consortium 等^[21]采用最小二乘法回归模型建立华法林剂量预测算法,来预测华法林的稳定维持剂量。

谷歌公司利用人们在网上的搜索记录和与流感密切相关的检索词条,建立了特定的系统和 4.5 亿个数学模型来预测流感的传播,甚至预测流感发生的地方。将得出的预测结果与美国 CDC 记录的实际流感病例进行对比后,他们的预测与官方数据的相关性高达 97%。这种预测就是建立在大数据基础上的,这是当今社会所独有的一种新型能力——以一种前所未有的方式,通过对海量数据进行分析,获得有巨大价值的产品、服务以及深刻的洞见^[22]。

3.3.3 个体化 大数据的差异性凸显医疗服务的个性化。基因测序是医疗服务个性化的代表,Bina

Technology 公司利用大数据来分析人类的基因序列,发现基因中罕见的病变信息。随着从基因测序中获得越来越多的遗传信息,这一技术将对我们的健康产生极大影响。

基因测序技术的不断发展,促进了个性化药物研发等新型的疾病治疗措施的出现。苹果公司总裁乔布斯在与癌症斗争的过程中,对自身所有 DNA 和肿瘤 DNA 进行排序,使得医生们能够基于他的特定基因组成,按所需效果用药。如果癌症病变导致药物失效,医生可以及时更换另一种药。通过个性化用药,将他的生命延长了好几年。

英国医疗保健局宣布将建立世界最大癌症患者数据库,为个性化的癌症治疗提供基础支撑。建立这个数据库的目的是推动“个性化医疗”,针对每位患者的癌症类别和具体情况对症用药。数据来自英国各地医疗机构的病例和 1 100 万份历史档案记录,并与威尔士、苏格兰和北爱尔兰的医疗保健数据库共享信息(姜国平. 大数据医疗实例分享. <http://chuansong.me/n/512148>.)。

3.3.4 共享性 数据共享是大数据应用的基石,与大数据的“四 V 特征”是相辅相成的。通过信息共享,连接各个信息孤岛,最大程度上增加数据量,为更多、更新的应用提供数据支撑。使用者可以接触到更多种类、更多时序的数据内容,为分析决策提供更加可靠的依据,大大加快信息流通速度,增加其时效性和可用性,同时产生更大的价值。

医学领域积累了海量的数据,但是大部分数据资源分散在不同国家、研究单位和研究者手中。我国于 2004 年 4 月正式启动“国家医药卫生科学数据共享工程”。包括 1 个网、6 个数据中心、40 个左右的主体数据库和 300 个左右的数据库(数据集系列),在这个框架中包含了多种不同层次的数据整合与资源组织方式。为政府卫生决策、医学科技创新、医疗保健、医学人才培养,以及全民健康提供数据资源和信息服务(共享国际生物医学数据. 科学时报. <http://www.bioon.com/biology/news/289239.shtml>.)。国际层面的医学数据共享不断发展,1997 年人类脑计划在美国正式启动,有 20 余家著名的研究所和大学参加。其目标是建立一个有关神经系统所有知识的全球管理系统和网络协同研究环境,使得有关脑的实验数据和研究成果能够灵活有效地被管理,从而最大限度地利用这些实验数据和结果,共享国际神经信息学资源,减少不必要的重复性研究和人力物力的浪费。

3.4 以数据库为基础的临床数据挖掘实例

临床数据库、电子病历与医学图像等半结构

化的数据一样,都是大数据在临床医学中的具体体现。与后两者不同的是,数据库采集信息的过程是有目的的、主动的,有专业人员进行信息的录入、整理并统一结构。因此,基于数据库所做的研究也更加的方便可行。数据库应用于临床工作是一种创新,临床研究不再仅仅局限于前瞻性的 RCTs,而是更加注重反映真实世界的情况,逐步从 RCTs 过渡到 BCT(Big-data Clinical Trial),可以预料到大数据时代的临床研究 BCT 将会取代 RCTs 成为主导的研究类型^[16]。

成立于 1989 年的美国胸外科协会(STS 数据库),至今已经涵盖了美国 95% 的心脏手术,收集了 500 万条手术记录^[23]。其中的先天性心脏手术(CHSD)数据库是 STS 数据库的重要组成部分,是北美最大的关注儿童先天性心脏畸形的数据库,被认为是医学专业临床结果数据库的金标准。近年来,基于 CHSD 数据库所进行的数据挖掘不断增加,大型数据库对提高医疗质量所起到的正向作用正在日益凸显。如 Welke 等^[24]基于 CHSD 数据库,探讨小儿心脏外科病例数量和死亡率之间的复杂关系;Pasquali 等^[25]基于 CHSD 数据库探讨新生儿 Blalock—taussig 分流术后的死亡率;Jacobs 等^[26]基于 CHSD 数据库采用多变量分析方法来研究病人术前因素的重要性;Dibardino 等^[27]基于 CHSD 数据库,采用多变量分析的方法来探讨性别和种族对进行先天性心脏手术结果的影响。

近年来,国内外心血管外科领域相继涌现出一系列高质量的临床数据库,如英国胸心外科协会(SCTS)数据库、澳大利亚和新西兰心胸外科数据库以及阜外心血管病医院的中国成人心外科数据库都极大的提高了心血管外科治疗的成功率。四川大学华西医院所建立的中国人心脏瓣膜置换术后抗凝治疗数据库,搜集了数万例心脏瓣膜置换术后患者有关抗凝治疗的住院及随访数据,为我国进行心脏瓣膜置换术后抗凝治疗的研究提供了坚实的数据支持^[28]。大型数据库的建立可以更好的支持临床数据挖掘工作,从而形成临床数据收集—挖掘—临床决策支持的闭环,达到医疗质量持续改进和提高的目的^[29]。

4 医学数据挖掘的趋势与展望

以大数据为基础的医学数据挖掘是一门涉及面广、技术难度大的新型交叉学科,是计算机技术、人工智能和现代医学相结合的产物,需要从事计算机、统计学的科研人员与广大医务工作者之间的通力合

作。数据挖掘具有自身独有的理念,给人们解决问题提供了一种新的思路和方法。因此,随着理论研究的深入和进一步的实践探索,医学数据挖掘必将在疾病的诊疗、医学科研与教学以及医院管理等方面发挥强有力的作用。

参考文献

- 1 李国杰. 大数据研究的科学价值. 中国计算机学会通讯, 2012, 8(9): 8-15.
- 2 Gantz J, Reinsel D. Extracting value from chaos. IDC iview, 2011, 1142 9-10.
- 3 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考. 中国科学院院刊, 2012, 27(6): 647-657.
- 4 潘璠. 大数据的概念并非无解. 青海统计, 2014, 10: 1-1.
- 5 Madden S. From databases to big data. IEEE Internet Computing, 2012, 3: 4-6.
- 6 Chen M, Mao SW, Liu YH. Big data: A survey. Mobile Networks and Applications, 2014, 19(2): 171-209.
- 7 刘维贵. “大数据”研究综述. 办公自动化: 综合月刊, 2014, 1: 27-30.
- 8 贺宪民, 吴聘, 于长春, 等. 数据挖掘技术在医学领域中的应用. 第二军医大学学报, 2004, 24(11): 1250-1252.
- 9 张永明. 数字图书馆个性化信息服务研究. 河南图书馆学刊, 2010, 2(3): 30-31.
- 10 夏虹. 探析数据挖掘在建筑企业管理中的应用. 城市建设理论(电子版), 2013, 13(9): 211-212.
- 11 Benson DA, Karsch-Mizrachi I, Lipman DJ, et al. GenBank. Nucl Acids Res, 2005, 33(suppl 1): D34-D38.
- 12 Srimani PK, Koti MS. A comparison of different learning models used in data mining for medical data; proceedings of the 2nd international conference on methods and models in science and technology. AIP Publishing, 2011, 1414(1): 51-55.
- 13 Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet, 2012, 13(6): 395-405.
- 14 Cios KJ, Moore GW. Uniqueness of medical data mining. Artif Intell Med, 2002, 26(1): 1-24.
- 15 Zhang Z. Big data and clinical research: focusing on the area of critical care medicine in mainland China. Quant Imag Med Surg, 2014, 4(5): 426.
- 16 Wang SD. Opportunities and challenges of clinical research in the big-data era: from RCT to BCT. J Thorac Dis, 2013, 5(6): 721-723.
- 17 Bath PA. Data mining in health and medical information. Annual review of information science and technology, 2004, 38(1): 331-369.
- 18 滕琪, 樊小毛, 何晨光, 等. 医疗大数据特征挖掘及重大突发疾病早期预警. 网络新媒体技术, 2014, (1): 50-54.
- 19 李颂. 大数据改变生活. 决策与信息, 2014, 374(10): 12-13.
- 20 刘娟. 基于数据挖掘的分类预测模型研究. 现代商业, 2014, (12): 47-47.
- 21 Consortium IWP, TEK, RBA, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med, 2009, 360(8): 753-764.
- 22 Naimi A, Westreich DJ. Big data: a revolution that will transform how we live, work, and think. Am J Epidemiol, 2014 April 23. [Epub ahead of print].
- 23 Saxena A, Newcomb AE, Dhurandhar V, et al. Application of

- clinical databases to contemporary cardiac surgery practice: where are we now? Heart, Lung and Circ, 2015, 19(2): S1443-9506.
- 24 Welke KF, O'Brien SM, Peterson ED, et al. The complex relationship between pediatric cardiac surgical case volumes and mortality rates in a national clinical database. J Thorac Cardiovasc Surg, 2009, 137(5): 1133-1140.
- 25 Pasquali SK, He X, Jacobs JP, et al. Evaluation of failure to rescue as a quality metric in pediatric heart surgery: an analysis of the STS congenital heart surgery database. Ann Thorac Surg, 2012, 94(2): 573-580.
- 26 Jacobs JP, O'Brien SM, Pasquali SK, et al. The importance of patient-specific preoperative factors: an analysis of the society of thoracic surgeons congenital heart surgery database. Ann Thorac Surg, 2014, 98(5): 1653-1659.
- 27 Dibardino DJ, Pasquali SK, Hirsch JC, et al. Effect of sex and race on outcome in patients undergoing congenital heart surgery: an analysis of the society of thoracic surgeons congenital heart surgery database. Ann Thorac Surg, 2012, 94(6): 2054-2060.
- 28 付博, 董力, 石应康. 中国人心脏瓣膜置换术后抗凝治疗数据库建设. 中国胸心血管外科临床杂志, 2013, 20(1): 2-9.
- 29 王剑, 张政波, 王卫东, 等. 基于重症监护数据库 MIMIC- 的临床数据挖掘研究. 中国医疗器械杂志, 2014, 38(6): 402-406.
- 收稿日期: 2015-04-22
编辑: 王唯、刘雪梅

引用本文: 秦文哲, 陈进, 董力. 大数据背景下医学数据挖掘的研究进展及应用. 中国胸心血管外科临床杂志, 2016, 23(1): 55-60. [Qin WZ, Chen J, Dong L. Progress and application of medical data mining under the background of big data. Chin J Clin Thorac Cardiovasc Surg, 2016, 23(1): 55-60.]