

· 论坛 / PERSPECTIVE ·

医疗健康大数据研究综述

颜延, 秦兴彬, 樊建平, 王磊*

中国科学院深圳先进技术研究院, 广东 深圳 518055

摘要: 本文从大数据的基本特点和医疗健康领域的大数据研究现状出发, 首先从大数据的数据来源和数据特点进行了综述, 概述了医疗信息数据和健康数据的采集方式; 然后列举了利用医疗健康数据挖掘进行健康信息数据分析的常用方法, 包括分类、回归分析、聚类、关联规则、特征分析、变化和偏差分析; 最后介绍了临床支持系统、远程医疗及远程监控、电子档案与公共健康、疾病模式分析几个方面的应用。

关键词: 大数据; 医疗健康; 数据挖掘

doi:10.11871/j.issn.1674-9480.2014.06.001

A Review of Big Data Research in Medicine & Healthcare

Yan Yan, Qin Xingbin, Fan Jianping, Wang Lei *

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China

Abstract: In this paper, the state-of-art of the big data research in medicine and healthcare is reviewed. The fundamental concepts of big data and its essential in the domain of healthcare and medicine were presented, the data sources and data characteristics were analyzed as well as the acquisition methods of medical data and healthcare information, the methods of data mining for medical & healthcare information were listed with applications, including classification, regression, clustering, association rules, feature analysis and deviation analysis method. At last, the application in clinical supporting system, telemedicine, electronic records with public healthcare and disease pattern analysis were explained.

Keywords: big data; medicine & healthcare; data mining

基金项目: 国家高技术研究发展计划(863计划)(2012AA02A604); 国家重点基础研究发展计划(973计划)(2010CB732606)

*通讯作者

1 概述

移动互联网、物联网、社交媒体网络相关的技术与应用,使全球范围内的数据容量正以前所未有的速度增长。互联网技术的高速发展,带来了数据技术、数据应用、数据价值的重大变革,以用户数据为基础,通过扩大数据的透明性、共享性和流动性,更多的相关关系被分析挖掘出来为各个研究领域提供了全新的思路。在欧洲粒子物理研究所的大型强子碰撞型加速装置的物理实验、临床实验中的病理机制研究等研究项目中,实验所得到的数据容量与维度也日益增加并且变得更加复杂^[1]。2011年5月,EMC公司在美国拉斯维加斯举办以“云计算相遇大数据”为主题的第11届EMC世界年度大会,大会正式提出了“大数据”(Big Data)的概念^[2]。随后IBM与麦肯锡等国外研究机构或咨询机构发布了“大数据”相关的研究报告,阐述了大数据的特征以及大数据为社会经济科学发展带来的机遇和对当前信息技术的挑战^[3-5]。在学术界,2008年《Nature》出版专刊Big Data,从互联网技术、网络经济学、超级计算、环境科学、生物医药等多个方面介绍了海量数据带来的挑战^[6];2011年《Science》推出关于数据处理的专刊Dealing with data,讨论了数据洪流所带来的挑战^[7];IEEE在2013召开Big Data国际会议并成立首届学术工作组,对大数据的理论、管理、信息安全、基础结构在互联网技术、医疗健康、环境科学等多个方面的应用和发展进行了深入探讨^[8-9]。

1.1 大数据特征

根据文献[10]的分类,大数据可以分为两种类型:一种是网络大数据,指在互联网上发生蕴含有丰富可被发掘的具有社会、商业或科研价值的大数据;另外一种类型,是指企业自身的产品和服务产生了大量的密集型“超大规模”或“海量数据”,通过对这些数据进行深入的挖掘分析,改进自身业务,改进后的业务吸引更多用户或客户,产生更大量的数据。

针对医院和健康服务机构来说,即在医院诊治病患以及健康服务机构采集用户生理信息过程中,

获得了大量的病历、诊断、筛查、检测数据,通过对这些数据进行挖掘分析,可以提升医院和健康服务机构的诊疗和服务水平。对健康服务机构来说,其移动健康服务或者移动医疗服务,获得的数据兼有上述两种特点。其数据具有Volume规模巨大,Velocity速度极快,Variety模态多样,Veracity真伪难辨的数据特点^[11]。

在医疗健康大数据方面,不仅包括在医院的诊疗过程中,常年积累下来的临床数据。在国内,医疗行业总体数据存储量并不大,对单个医疗机构而言,以1TB~50TB为主,也有个别信息化水平较高的机构达到PB级别。根据医疗数据的保存要求,门诊记录保存时间不得少于15年,住院病历保存时间要求大约为30年,期间所积累的数据样本量和数据维度都具有相当大的可分析价值。在国内外的医疗机构及医疗研究机构中,从临床实验中的批量采集数据积累而成的数据量级也是巨大的,例如HCUP项目建立了大约1亿观测人群的医疗和健康及其消费数据库,890394名退役军人的2002-2006的糖尿病治疗跟踪研究等;在临床治疗、预后和趋势研究、疾病风险因子提取和新型治疗技术的开发方面,随着现代检测、存储技术的发展,采集到的数据的复杂度和数据容量都在不断增大。

1.2 医疗信息数据

医院信息管理系统,是以计算机为基础,用以简化所有医院医疗和管理信息,是医疗大数据发展的来源和应用领域。随着信息技术的发展,医院信息系统得到快速发展,根据2005年CHIMA医院信息化调查资料数据,在我国许多医院临床信息系统已经得到快速的应用和发展,LIS系统占39.14%,住院医生工作站系统占35.04%,门诊医生工作站系统占32.99%,在沿海经济发达地区的信息化突出的医院,电子病历、全院PACS、移动、无线、PDA、Tablet PC、RFID、万兆网络、服务器集群等先进的系统和先进的IT技术已经开始应用^[12]。

然而大数据的技术与模式发展在目前的医疗信息化问题上,存在着诸多问题。首先是信息缺乏与信

息的不对称：医学信息的缺乏与医学信息的不对称性，循证医学能提供证据的也仅能解决小部分问题，其余的临床问题的解决还有赖于未来高质量、大规模的临床研究和观察结果。从数据科学的角度来看，这种医学信息的不对称性在实现数据挖掘以及统计推断中，其后果是致命的。再者是医生与医学信息的不对称性，在许多的基层机构中，某些疾病的诊断方式方法还停留在较为原始的水准，不能够获得和使用最新的医疗技术，其带来的数据不可靠性也会带来错误数据的积累。同时，行业复杂、标准化存在巨大困难：患者的个体差异，疾病的种类繁多，疾病临床现象与生理检测参数关系复杂，使得现有的理论和技术手段很难完成医疗信息的标准化和自动化；除此以外，随着医学理论与技术的发展，新的医学检查手段、诊断标准、治疗方法不断发展以及新的疾病不断产生与变化，医疗信息的发展还存在未知的领域。

因此，医疗信息的数据挖掘以及分析与互联网的大数据技术的区别在于：互联网中，用户的行为是确定的，比如用户的浏览行为、浏览习惯，都具有强烈的数据确定性；而在医疗信息中，诚然在一些生理数据的检查，比如心电、血压、生化等参数信息是确定的，但然后其诊断、临床记录存在着部分不确定因素，其带来的对数据挖掘和数据分析的影响是不容忽视的。

1.3 移动健康数据

在传统的眼光看来，“健康”与“医疗”处于不同的领域，有一定的关联但是从未有目前这样的紧密的联系。大数据革命带来的是两个领域的相通与结合。一个理想的愿景是：由“可穿戴式设备”或者其他采集终端，搜集到用户的生理数据，存储到云端并结合病历数据，完成数据的分析与处理，医生查看相关结果，给出相应的诊断或者康复建议。这样的系统可以在每时每刻进行，大体的系统结构是在用户连续采集生理数据，服务中心（或者云端）实现数据的分析，然后推送健康指导或者健康建议，实例有日常的健康监督、运动健身指导，高血压、糖尿病的日常管理等等。然而这种系统离参数监测和医疗记录数据的

融合及智能分析的设想相去甚远。

1.4 本文内容

大数据的应用是医疗健康信息化智能化的最新趋势，首先对各种医疗信息系统和移动健康的相关概念进行了梳理，然后对医疗健康大数据系统结构和相关智能分析技术的发展现状进行了分析，最后总结了大数据相关分析技术在医疗健康中的运用。

2 医疗健康大数据的数据采集

大量的数据可以分析出疾病、症状及实验室数据的相关性，从而帮助临床研究人员建立针对某一些典型疾病的预测模型。在医院的诊疗过程中，针对各个科室的特定应用，积累了长期的与特定疾病相关的临床监测参数，并随着医院的运营过程得到了大量的数据的积累。

同时，随着移动互联网技术和穿戴式医疗设备及技术的发展，通过各种穿戴式设备所获取的用户生命体征，为用户健康数据的获取提供了极大的便利。一方面，可以通过对这些健康数据进行分析获取用户的健康信息以指导运动、饮食等生活习性；另一方面，与医疗数据的结合可以提高用户疾病诊断的科学性和诊断精度。

2.1 医院信息系统

医院信息系统（Hospital Information System），公认的定义为 Collen 提出的：利用电子计算机和通讯设备，为医院所属各个部门提供病人诊疗信息和行政管理信息的采集、存储、处理、提取和数据交换的能力并满足授权用户的功能需求的平台^[13]。医院信息系统包括医学影像信息系统、临床信息系统、放射学信息系统 RIS、实验室信息系统 LIS 等，除此之外，医院信息系统的主要部分为门诊数据和病房数据，也是医疗大数据分析的重要组成部分。

医学影像存档与通讯系统（PACS）包含了医院影像科室的核磁检查数据、CT 图像数据、超声图像数据、X 光检测数据的海量存储,包括设备本身具有的

辅助诊断和管理数据。

临床信息系统 (CIS) 的主要目标是支持医院医护人员的临床活动, 采集和处理病人的临床医疗信息, 丰富和积累临床医学知识, 并提供临床咨询、辅助诊疗、辅助临床决策, 提高医护人员工作效率和诊疗质量^[14]。比如医嘱处理系统、病人床边系统、重症监护系统、移动输液系统、合理用药监测系统、医生工作站系统、实验室检验信息系统、药物咨询系统等均属于临床信息系统范围。

实验室信息系统 (LIS), 其主要功能是讲实验仪器的检验数据经过分析后, 生成检验报告, 存储在网络数据库中, 医生可以通过数据库查看检验标本的检验结果辅助临床治疗^[15]。

电子病历系统, 与上述系统不同, 电子病历以病人人为中心, 将病人诊断过程中产生的诊疗数据和检查数据集合为具有统一形式的记录, 是最具有价值的数据来源^[16], 也是真正可以与长期健康监测整合为统一数据样本的医疗数据信息。以计算机化的病历系统或者基于计算机的病人记录用以集成患者医疗信息, 可以为教学、科研和决策提供资料来源。电子病历的发展目标主要是加速患者医疗信息流通, 使患者信息在医疗系统内随时随处可以得到, 提供纸张病历无法提供的服务, 在信息化的角度来讲, 方便了大数据挖掘中的信息提取, 标准化的电子病历有效的提高决策效率和医疗服务质量。

2.2 医疗大数据的数据特点

传统医疗行业中, 医院信息系统完成了医院内部的流程控制、数据积累等工作。医疗行业早就遇到了海量数据和非结构化数据的挑战, 而近年来很多国家都在积极推进医疗信息化发展, 这使得很多医疗机构有资金来做大数据分析^[17]。医疗数据是医疗人员对病人诊疗过程中产生的数据, 包括病人的基本情况、行为数据、诊疗数据、管理数据、检查数据、电子病历等。现代医院中将上述数据存储于医院的各个信息系统之中, 是医疗大数据分析的基础。

按照大数据的概念来看, 医疗数据的 Volume、Variety、Value、Velocity 四个特征都是显而易见的,

除此之外, 医疗大数据具有多态性、不完整性、时效性、冗余性、隐私性等特点^[18]。

多态性: 医疗数据的表达格式包括文本型、数字型和图像型。文本型数据包括人口特征、医嘱、药物使用、临床症状描述等数据; 数字型数据包括检验科的生理数据、生化数据、生命体征数据等; 图像型数据包括医院中的各种影像学检查如 B 超、CT、MRI、X 光等图像资料。在文本型数据中, 数据的表达很难标准化, 对病例状态的描述具有主观性, 没有统一的标准和要求, 甚至对临床数据的解释都是使用非结构化的语言。多态性是医学数据区别于其他领域数据的最根本和最显著的特性。这种特性也在一定程度上加大了医疗数据的分析难度和速度。

不完整性: 医疗数据的搜集和处理过程存在脱节, 医疗数据库对疾病信息的反映有限。同时, 人工记录的数据会存在数据的偏差与残缺, 数据的表达、记录有主观上的不确定性。同一种疾病并不可能全面由医学数据反映出来, 因此疾病的临床治疗方案并不能通过对数据的分析和挖掘而得出。另外, 从长期来看, 随着治疗手段和技术手段的发展, 新类型的医疗数据被创造出来, 数据挖掘的对象的维度是在不停的增长的^[18]。

时效性: 病人的就诊、疾病的发病过程在时间上有一个进度, 医学检测的波形信号 (比如说心电图) 和图像信号 (MRI、CT 等) 属于时间函数, 具有时效性。例如心电图信号检测中, 短时的心电无法检出某些阵发性信号, 而只能通过长期监测的方式实现心脏状态的监测^[19]。

冗余性: 医疗数据中存在大量的相同或类似信息被记录下来。比如常见疾病的描述信息, 与病理特征无关的检查信息。

隐私性: 在对医疗数据的数据挖掘中, 不可避免的会涉及到患者的隐私信息, 这些隐私信息的泄露会对患者的生活造成不良的影响。特别是在移动健康和医疗服务的体系中, 将医疗数据和移动健康监测甚至一些网络行为、社交信息整合到一起的时候, 医疗数据的隐私泄露带来的危害将更加严重。大数据分析中隐私保护要注意两个方面: 其一, 用户身份、姓名、

地址和疾病等敏感信息的保密；其二，经分析后所得的私人信息的保密^[20]。

2.3 健康大数据的采集与特点

移动设备的性能快速提升和无线网络的广泛覆盖，以及穿戴式设备与技术的发展和移动应用的广泛推广，为健康服务和个人健康管理创造了巨大的空间。其中移动健康服务是最常被提及到的概念，移动健康（mobile health, mHealth）是将通信技术应用于卫生环保领域，提供健康服务和信息，实现“健康传感终端+移动通信平台+健康服务”，从而提供实时、连续、长期的健康服务^[21]。移动健康不仅可以实现普适意义上的健康连续监测，作为用户的医疗诊治和监护参考，而且在医疗资源短缺的情况下，可以改善医疗卫生服务的整体水平。

健康服务区别于传统医疗服务的主要特点在于，传统的医疗信息系统和信息服务，适用于单个医院或者是区域内部，其工作原理更趋向于一个局域网的内部的信息共享和信息服务系统。而移动健康的重要进步在于将移动通信技术、互联网技术和健康服务融合起来，通过智能手机、平板电脑等移动设备终端，通过无线通信网络接入到互联网之中^[21]。

移动健康的发展原型是远程医疗（telemedicine, telehealthcare），即利用远程通信技术、全息影像技术、新电子技术和计算机多媒体技术发挥大型医学中心医疗技术和设备优势对医疗卫生条件较差的及特殊环境提供远距离医学信息和服务^[22-23]。互联网大数据的兴起和穿戴式医疗设备，在健康服务上迈向了新的一步，建立在互联网的高速通信技术之上的数据获取和交互与人的联系更加紧密，通过用户近场的生物传感器和移动 APP 获得了大量的健康信息（体重、心率、运动信息、血压、血氧），同时伴随着的是相关的用户信息（医疗记录、语音、视频、地理位置）。移动健康真正可以实现用户的随时、随地、随身获得相关的健康信息。

2.4 移动健康大数据的采集终端

移动终端采集到的数据与医疗机构通过专业设

备采集到的数据具有不同之处。首先，由于移动终端设备的功能和技术限制，获取的健康数据无法达到专用的医疗设备所能采集到的设备的水准，在专业性和全面性上都无法与医疗机构的检测相比。再者，移动健康所具备的随时、随地性是优于医疗机构诊治信息的地方。最后，移动互联网中的一些新的非结构化数据，比如说社交文本信息、非医学图像信息、语音信息等，在新的信号分析和数据分析技术下，也可以为传统的医疗诊断带来新的思路。现有的穿戴式传感器和移动终端可获得的数据和相关技术如下：

(1) 心电数据

心电图是反映心脏兴奋的电活动过程，它对心脏基本功能及其病理研究方面，具有重要的参考价值。心电图可以分析与鉴别各种心律失常；也可以反映心肌受损的程度和发展过程和心房、心室的功能结构情况^[24]。在日常生活中对患者进行心电监护可以为医生临床诊断提供参考，对普通人而言，心电图有助于用户监测身体健康状态。在实现移动健康的心电信号监测中，与 Holter 系统、TTM 心电监护系统有所区别，其具有的移动通信功能可以为用户提供更大活动范围、更为灵活的通讯方式^[25]。传统的心电图机的心电测试可以以居家的方式在用户端实现，用户只需经过简单的操作就可以完成心电信号的采集^[26]。心电信号的长期监测也在传统 Holter 之上实现了终端智能设备上的应用^[27]，更低功耗^[28]和更长续航时间^[29]的心电监测设备也得到了相应的应用。心电参数的监测被广泛应用到患者疾病跟踪^[30]、运动员生理状态监测^[31]中。心率参数的监测被广泛的应用到一些运动相关的移动应用之中^[32]。

(2) 生命体征参数

生命四大体征包括呼吸、体温、脉搏、血压，医学上称为四大体征。它们是维持机体正常活动的支柱，缺一不可，不论哪项异常也会导致严重或致命的疾病，同时某些疾病也可导致这四大体征的变化或恶化，用户生命体征数据的采集对用户疾病预防及治疗跟踪具有重要的意义。穿戴式设备以及智能终端可以通过集成的生物传感器实现对生命体征参数的采集。

文献[33]对比了几种穿戴式呼吸率监测的方法的准确性和可靠性。文献[34]介绍了一种基于声音的穿戴式呼吸率监测方法。文献[35]介绍了一种电容传感器用以完成呼吸率的监测。文献[36-38]介绍了可穿戴连续血压监测方法。文献[39]介绍了一种腕式血氧监测装置在睡眠监测中的应用。在专用的穿戴式设备之外,利用移动终端的图像采集功能,扩展出一些基于图像处理技术的生命体征监测技术,文献[40-41]对利用摄像头实现生命体征参数的监测技术进行了介绍。文献[42]系统总结了穿戴式传感器在移动健康监护和诊疗中的应用。

(3) 运动健康

伴随移动互联网快速发展,运动健康类的移动应用以其关注程度和实现便利性得到了最为广泛的推广。利用移动终端的定位、记录和交互式的引导功能,用户的健康数据、个人信息得到了有效的积累。文献[43]介绍了通过网络健身平台及外设健身产品,实现运动健身的网络互动的网络健身概念和网络健身干预体系。文献[44]描述了通过运动传感器的数据监测来实现对人的行为和运动进行记录应用在慢性病管理中。在运动健康的健康数据采集,主要分为两种数据类型:一种是基于用户的行为模式以及活动记录相关的数据,主要集中在地理信息、锻炼信息、习惯记录等,部分会采用传感器用以辅助完成;另外一种是基于运动传感器的身体运动信息的监测,文献[45]为代表的基于传感器的行为状态检测,应用于用户的跌倒检测^[45]、步态分析^[46]等基于人体传感器网络的健康监护系统之中。

(4) 其他

实际上,现有的健康监测技术往往是多参数的同步提取,比如在基于图像信息的生理参数提取技术中,可以通过 PPG 图像完成脉率、呼吸率的提取^[41]。又如在基于织物的生理参数采集系统中,可以完成心电、体温、呼吸、运动等多参数的采集^[47]。基于多参数的传感系统传感器技术的发展,带来的是对人体参数的检测和记录更加全面,对个人的健康和状态的分析与认识也更加的清楚,但是所带来的数据类型也越来越多,数据的结构趋于更加复杂。

另外,在上述内容中提及的健康参数之外,还有一些未被归类的相关信息,例如针对糖尿病人的血糖分析参数^[48],针对运动员的肌电信号^[49],脑电信号监测等。诸多的生理信号、参数都是在医生的指导下完成的,移动互联网技术为这些信号的采集、存储带来了相应的便利,但本文仍将远程监护归纳到医疗记录或理疗数据范畴。相关的数据或者记录从宏观角度来看来应当蕴含在医院信息系统之中。

3 医疗健康大数据处理方法

医疗健康数据是持续、高增长的复杂数据,蕴含的信息价值也是丰富多样的,对医疗健康数据的有效存储、处理、查询和分析,挖掘其潜在价值,发现医学知识,将深切影响人类健康水平和治疗手段。在传统的医学统计方法的基础上,新的模型与技术的出现,为从数据中获取新知识提供了新的思路。

医疗健康数据挖掘进行健康信息数据分析常用的方法包括分类、回归分析、聚类、关联规则、特征分析、变化和偏差分析。针对不同的类型的病人对不同类型的生理数据、健康感知数据进行推理判断,大数据分析技术实现了服务临床治疗、预测疾病发病情况、跟踪病人病情等目的。下面将介绍在医疗健康感知领域上述大数据方法的应用现状。

3.1 分类

分类是找出数据库中一组数据对象的共同特点并按照分类模式将其划分为不同的类,其目的是通过分类模型,将数据库中的数据项映射到某个给定的类别^[50]。其实例可以应用到用户的分类、用户的属性和特征分析、用户治疗效果评价等。文献[51]利用了病人的人口学信息、对情绪图片库的生物反馈信息、自主神经的生理学特性信息作为输入特征,使用决策树算法用以对生物唤起讯号分类。文献[52]对比了基于规则、基于决策树、基于人工神经网络的糖尿病人的健康数据的模式分类方法。文献[53]提出了一种基于用户医疗和健康监护数据的医护干预分类方法。

除了对医疗健康数据库的数据分类之外, 对生理信号进行模式分类, 建立生理信号的正常异常模型, 为临床诊断提供依据。分类的精确性和模型的准确度随着数据的积累得到了提高, 例如动态心电图中的异常波形检测^[54]、波形分类^[55]算法等。医疗数据的自动分类和判别提高了计算机辅助诊断水平, 为医生的诊疗提供了方便。

3.2 回归分析

回归分析方法反映的是事务数据库中属性值在时间上的特征, 产生一个将数据项映射到一个实值预测变量的函数, 发现变量或属性间的依赖关系, 其主要研究问题包括数据序列的趋势特征、数据序列的预测以及数据间的相关关系等^[56]。例如对医院信息系统的医疗风险因素的回归分析, 分析各个影响因素与医疗风险之间的联系及引起风险的概率变化, 用以指导医院的风险管理^[57]。临床心理治疗中通过对实验结果的回归分析, 研究人员可以高效地区分有效预测和无效预测, 并且发现预测变量之间的联系, 并对临床的治疗决策提供预测模型^[58]。文献[59]使用回归模型对父母调查报告、孩子调查报告、生物标记、行为与儿童焦虑抑郁建模, 并将模型应用到临床指导。

3.3 聚类

聚类分析是把一组数据按照相似性和差异性分为几个类别, 其目的是使得属于同一类别的数据间的相似性尽可能大, 不同类别中的数据间的相似性尽可能小^[60]。文献[61]通过对青少年和成年人的酗酒成瘾状况及心理测试的结果聚类分析来完成酗酒人格模式的知识发现。文献[62]通过对疼痛反应结果的聚类分析, 完成了对热性疼痛、压力性疼痛、缺血性疼痛的诱因分析。文献[63]对聚类方法在心理健康中的应用做了综述。聚类方法的一般思路是通过观测数据样本的亲疏关系的统计量, 根据某种准则使同类型的数据差别较小, 类与类之间差别较大, 以此完成个体或者变量的区分。在医疗健康记录的关键词分类^[64]、生理信号分析^[65-66]中发挥了重要作用。

3.4 关联规则

关联规则是描述数据库中数据项之间所存在的关系的规则, 即根据一个事务中某些项的出现可导出另一些项在同一事务中也出现, 即隐藏在数据间的关联或相互关系。在电子健康档案中, 大量用户的个人信息、健康信息、临床诊疗信息等, 可以应用到疾病的检测和预测中。文献[67]通过对比房颤病人的医疗数据研究房颤与脑梗塞的关联规则。文献[68]对门诊病人的医疗档案和异常的筛查结果进行了关联规则分析。文献[69]提出了一种基于医院信息系统和医疗记录的临床推荐系统。医疗记录中的关联规则发现, 有助于新的医学知识的发现和发病风险分析^[70]。

3.5 特征分析

特征分析是从数据中的一组数据提取出关于这些数据的特征式, 这些特征式表达了该数据集的总体特征。在长期个人健康相关的生理信号监测中, 通过对采集数据的特征分析, 可以有针对性的完善病人治疗方案。文献[71-72]介绍了在移动健康中对长期监测的心电信号的特征提取。文献[73]介绍了慢性癫痫症的脑电监测信号的模式提取。文献[74]介绍肌电信号特征提取的系统与应用。文献[75]介绍了老年痴呆症临床应用中的头部核磁图像特征提取。无论是在临床生理信号, 还是在影像检测中, 大样本量的数据特征提取技术得到广泛的应用, 为更好的自动检测方法和临床治疗水平的提高提供了技术支撑。

3.6 小结

关于大数据分析技术在临床和健康监测系统中的应用, 相关的综述对数据挖掘或者机器学习都有详细的介绍^[76-78]。对医疗信息和健康数据的智能分析不停的向前发展, 医疗健康领域内的人工智能将随着大数据时代的发展而不停更新其内容, 随着新的计算技术、计算平台出现和计算能力的提高, 更深层次的规律和知识将得以发现, 将极大的提高临床诊疗和医疗服务水平。

4 医疗健康大数据应用

医疗行业的传统数据应用具有重要的参考价值, 必须明确的是大数据的发展是建立在已有的技术基础、数据积累之上的拓展。新的信息分析技术和通讯技术为传统的医疗网络应用和数据分析带来了新的思路。

在对用户的诊疗数据、健康监测数据的采集和分析的基础之上, 可以实现用户身体状况的预测、监控, 甚至可以确定用户是哪一类的疾病的易感人群。提高用户的健康状况水平, 降低用户的患病风险。精准分析包括病人体征数据、费用数据和疗效数据在内的大型数据集, 可以帮助医生确定临床上最有效和最具有成本效益的治疗方法。医疗护理系统将有可能减少过度治疗, 比如避免副作用大于疗效的治疗方式。

4.1 临床决策支持系统

临床决策支持, 是指医生在诊疗过程中, 能对医生的实时诊疗决策制定做出帮助的各种资源。常见的有科研文献、在线期刊、专家会诊意见、循证医学证据、临床决策支持系统(CDSS)等。临床决策支持系统, 是通过数据、模型等, 以人机交互辅助临床工作人员决策的计算机应用系统。

得益于对非结构化数据的分析能力的日益加强, 临床决策支持系统在大数据分析技术的帮助下变得更加智能。比如可以使用图像分析和识别技术, 识别医疗影像数据, 或者挖掘医疗文献数据建立医疗专家数据库, 从而为医生提出诊疗建议。文献[79]介绍一种基于生理数据的云计算用药决策支持系统, 使用了基于生理数据和药物剂量及临床表现的历史数据用以指导早产儿药物剂量。文献[80]探讨了一些不便到达的特殊地区的远程临床咨询服务系统的技术挑战, 介绍了基于移动终端的远程临床决策支持系统。文献[81]论述了利用临床心脏影像大数据支持的人工智能和先进计算, 用以实现个性化的治疗。利用机器学习对临床数据建模, 用以实现疾病的预测、康复和临床决策支持, 为医生的治疗提供了新的思路^[82-84]。

4.2 远程医疗及远程病人监控

从对慢性病病人的远程监控系统采集数据, 并将分析结果反馈给监控设备(查看病人是否正在遵从医嘱), 从而确定今后的用药和治疗方案。利用移动智能终端及穿戴式设备实现对病人的远程监控, 将患者的状态、参数纳入到病人的病历之中。特别是在慢性病患者治疗过程中, 远程监护可以有效的监测用户的健康状况。文献[85]介绍的一种针对糖尿病病人的远程医疗和远程病人监护系统, 系统包括可穿戴式的、便携式的生理参数监测设备, 可以对病人的生理参数进行实时监控, 在专家系统和医生的监控之下进行长期的跟踪治疗, 是远程监控的典型应用。文献[86-88]介绍了移动健康的相关技术和体系结构, 并对互联网通讯技术的发展对远程医疗的推动进行了详细的阐述。

4.3 电子档案分析与公共健康

在病人档案方面应用高级分析可以确定哪些人是某类疾病的易感人群, 进行药物使用的安全性分析。文献[89]通过对相关病人的电子病历以及药品代理商的药物资料进行数据分析, 用以完成药品安全监测, 防止药品滥用事件的发生。文献[90]综述了几种通过对基因和遗传等数据进行建模来推断疾病的潜伏期、易感群体、传染性的方法。通过全面分析病人特征数据和疗效数据, 然后比较多种干预措施的有效性, 可以找到针对特定病人的最佳治疗途径。文献[91]介绍了通过电子医疗记录进行公共卫生监测和传染病控制的几种方法, 文献[92]以心衰为例子综述了使用电子病历进行疾病预测建模的挑战与策略, 并对几种智能算法进行了分析。文献[93]对使用电子档案进行数据挖掘完成研究工作和临床治疗的意义和面临的挑战进行了阐述, 电子病历的数据挖掘有利于新的疾病分级策略的建立和对未知疾病的相关临床症状进行分析, 结合基因数据, 可以实现对基因表达的生物机制研究。

4.4 疾病模式分析与个性化医疗

美国罗彻斯特大学医学院精神病学和内科教授恩格尔(G. L. Engel)在1977年科学杂志上发表的题为

“需要新的医学模式, 对生物医学的挑战”的文章^[94]中指出, 现有的占统治地位的疾病模式是生物医学模式, 以分子生物学的可测量的生物学变量来分析疾病, 没有将病患的社会、心理和行为方面纳入到医学模式之中。通过对病人生理参数的长期监测, 挖掘病人电子档案, 实现疾病的预测、疾病的建模已经广泛应用在医疗领域^[95-98]。大数据背景下, 把患者的健康数据包括锻炼习惯、生活习惯、社交媒体信息等等纳入到疾病模式的分析和建模中来, 可以更有针对性的针对个体实现个性化的治疗, 也是生物心理社会医学的一个发展方向, 例如可以通过对社交媒体数据进行文字关键字分析来分析青少年心理压力^[99]。

5 总结与展望

本文从大数据的基本特点和医疗健康领域的大数据发展出发, 分析了医疗健康大数据的数据来源和数据特点, 综述了基于医疗大数据和健康大数据的应用, 最后对医疗健康大数据应用中使用的大数据方法进行了概述。

首先, 医疗领域的海量数据的积累, 并不是完全的新概念, 而是在医疗机构的临床治疗、实验中一直存在的, 随着信息化程度的加深, 越来越多的诊疗数据以可分析的方式逐渐积累。同时, 随着移动互联网和穿戴式医疗的发展, 用户的日常监测变得可行, 获取的数据方式与数据来源得到了极大拓展。

然后, 如何在海量的数据基础上获得有价值的信息以及如何使用新的方法来完善现有的诊疗知识库, 不断更新数据挖掘方法起到了非常重要的作用。分类、回归、聚类等数据挖掘方法在医疗健康大数据分析中获得的分析结果在临床决策支持、远程医疗与远程病人监控、电子档案分析与公共健康方面, 都得到了广泛的应用。

总之, 医疗与健康与人类的生活息息相关, 随着技术的发展, 如何更好的利用技术服务人类, 促进人类的发展, 在大数据时代背景下变得更加迫切。医疗健康大数据的应用, 不仅仅可以为人类带来更好的医疗健康服务, 更为重要的是在应用中, 大数据方法可

以不断发现新的知识内容, 促进医学知识和医学技术的进步。

参考文献

- [1] LYNCH C. Big data: How do your data grow? [J]. Nature, 2008, 455(7209): 28-29.
- [2] 黄亮. 解读大数据和 CDO: 概念大过实际意义? [J]. 网络与信息, 2013, 26(9): 24-26.
- [3] JACOBS, ADAM. The pathologies of big data [D]. Communications of the ACM, 2009, 52(8): 36-44.
- [4] BUGHIN J, CHUI M, MANYIKA J. Clouds, big data, and smart assets: Ten tech-enabled business trends to watch [D]. McKinsey Quarterly, 2011, 56(1): 75-86.
- [5] BROWN B, CHUI M, MANYIKA J. Are you ready for the era of "big data" [D]. McKinsey Quarterly, 2011, 4: 24-35.
- [6] BUXTON B, HAYWARD V, PEARSON I. Big data: the next Google [J]. Nature, 2008, 455(7209), 8-9.
- [7] OVERPECK J T, MEEHL G A., BONY S. Climate data challenges in the 21st century [J]. Science, 2011, 331 (6018): 700-702.
- [8] ZHENG Z, ZHU J, LYU M R. Service-generated Big Data and Big Data-as-a-Service: An Overview. IEEE International Congress on Big Data, 2013, 403-410.
- [9] 黄哲学, 曹付元, 李俊杰. 面向大数据的海云数据系统关键技术研究[J]. 网络新媒体技术, 2012, 1(6):20-26.
- [10] 何非, 何克清. 大数据及其科学问题与方法的探讨[J]. 武汉大学学报: 理学版, 2014, 1: 1-12.
- [11] GUPTA R, GUPTA H, MOHANIA M. Cloud Computing and Big Data Analytics: What Is New from Databases Perspective? [J]. Springer Berlin Heidelberg In Big Data Analytics, 2012, 7678:42-61.
- [12] 中国医院协会信息管理专业委员会. 中国医院信息化发展研究报告 (白皮书) [J]. 中国数字医学, 2008, (6): 11-19.
- [13] COLLEN M F. A brief historical overview of hospital information system evolution in the United States [J].

- International journal of bio-medical computing, 1991, 29(3):169-189.
- [14] 薛万国,李包罗. 临床信息系统与电子病历[J]. 中国护理管理, 2009, 2: 77-80.
- [15] 杨大千,徐根云,朱阳军. 临床实验室信息系统的研究与开发[J]. 中国医疗器械杂志, 2005, 29(1): 23-26.
- [16] AMARASINGHAM, RUBEN, BILLY J. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Medical care, 2010, 48(11): 981-988.
- [17] MURDOCH, TRAVIS B, ALLAN S. The inevitable application of big data to health care. JAMA, 2013, 309(13): 1351-1352.
- [18] 陈功, 范晓薇, 蒋萌. 数据挖掘与医学数据资源开发利用[J]. 北京生物医学工程, 2010, 29(3): 323-328.
- [19] PARTA A,FAES L,MASE M. An integrated approach based on uniform quantization for the evaluation of complexity of short-term heart period variability: application to 24h Holter recordings in healthy and heart failure humans [J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2007,17(1): 015117.
- [20] 胡新平. 医疗数据挖掘中的隐私保护 [J]. 医学信息学杂志, 2009, 8: 1-4.
- [21] 姚志洪. 跨入移动健康时代 [J]. 中国医疗器械信息, 2013, 18(11): 1-7.
- [22] PEREDNIA, DOUGLAS A, ALLEN A. Telemedicine technology and clinical applications [J]. JAMA, 1995, 273(6): 483-488.
- [23] WOOTTON, RICHARD, JOHN C. Introduction to telemedicine [D]. Royal Society of Medicine Press, 2006.
- [24] 庞志茹. 动态心电图与常规心电图诊断冠心病患者心律失常的比较[J]. 中外医疗, 2009, 13: 1-2.
- [25] 朱凌云. 移动心电监护系统ECG信号的智能检测与分析方法研究 [D]. 重庆大学, 2003.
- [26] YAN Y, QI L, YAN L. A home-based health information acquisition system [J]. Health Information Science and Systems, 2013, 1:1-12.
- [27] ORESKO, JOSEPH J, JIN Z. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing [J]. IEEE Transactions on Information Technology in Biomedicine, 2010, 14(3): 734-740.
- [28] PARK C, CHOU P H, BAI Y. An ultra-wearable, wireless, low power ECG monitoring system. Biomedical Circuits and Systems Conference. 2006.
- [29] HIGGINS S L. A novel patch for heart rhythm monitoring: is the Holter monitor obsolete? [J]. Future cardiology, 2013, 9(3): 325-333.
- [30] YOO J, YAN L, LEE S. A wearable ECG acquisition system with compact planar-fashionable circuit board-based shirt [J]. IEEE Transactions on Information Technology in Biomedicine, , 2009,13(6): 897-902.
- [31] GARGIULO G, BIFULCO P, CESARELLI M. An ultra-high input impedance ECG amplifier for long-term monitoring of athletes. Medical devices (Auckland, NZ), 2010, 3: 1-9.
- [32] PLEWS D J, LAURSEN P B, STANLEY J. Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring. Sports Medicine, 2013, 43(9): 773-781.
- [33] LANATA A, SCILINGO E P, NARDINI E. Comparative evaluation of susceptibility to motion artifact in different wearable systems for monitoring respiratory rate. IEEE Transactions on Information Technology in Biomedicine, 2010, 14(2): 378-386.
- [34] ZHANG J, SER W, GOH D Y. A novel respiratory rate estimation method for sound-based wearable monitoring systems. In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (pp. 3213-3216). IEEE.
- [35] KUNDU, KUMAGAI S, SASAKI M. A Wearable Capacitive Sensor for Monitoring Human Respiratory Rate. Japanese Journal of Applied Physics, 52(4S), 04CL05.
- [36] FERREIRA M, DAVID M D. A real time, wearable ECG and blood pressure monitoring system. In Information

- Systems and Technologies (CISTI), 2011 6th Iberian Conference on (pp. 1-4).
- [37] FRANCO J, AEDO J, RIVERA F. Continuous, Non-Invasive and Cuff-Free Blood Pressure Monitoring System. In Andean Region International Conference (ANDESCON), 2012, VI: 31-34
- [38] SHALTIS, PHILIP A, ANDREW R. Wearable, cuff-less PPG-based blood pressure monitor with novel height sensor. In Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE (pp. 908-911). IEEE.
- [39] ROFOUEI, MAHSAN, MIKE S. A non-invasive wearable neck-cuff system for real-time sleep monitoring. In Body Sensor Networks (BSN), 2011 International Conference on (pp. 156-161). IEEE.
- [40] SCULLY C, LEE J, MEYER J. Physiological parameter monitoring from optical recordings with a mobile phone. IEEE Transactions on Biomedical Engineering, 2012, 59(2): 303-306.
- [41] ASADA H, SHALTIS P, REISNER P. Mobile monitoring with wearable photoplethysmographic biosensors. IEEE Engineering in Medicine and Biology Magazine, 2005, 22(3): 28-40.
- [42] PANTELOPOULOS, ALEXANDROS, NIKOLAOS G. A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Transactions on Applications and Reviews, 2010, 40(1): 1-12.
- [43] 陆大江. 基于云健康的生活方式远程监控[J]. 中国医疗器械信息, 2013, 18(11): 8-17.
- [44] PATEL, SHYAMAL, HYUNG P.. A review of wearable sensors and systems with application in rehabilitation. Journal of neuroengineering and rehabilitation, 2012, 9(1): 21.
- [45] ZHANG, TONG, WANG. Fall detection by wearable sensor and one-class SVM algorithm. Lecture Notes in Control and Information Science, Springer Berlin Heidelberg, 2006, 345: 858-863.
- [46] LIU, TAO, YOSHIO I. Development of a wearable sensor system for quantitative gait analysis. Measurement, 2009, 42(7): 978-988.
- [47] DI R, MARCRO, FRANCESCO R. MagIC system: A new textile-based wearable device for biological signal monitoring. Applicability in daily life and clinical setting. In Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the (pp. 7167-7169). IEEE.
- [48] SONKSEN P H, JUDD S L, LOWY C. Home monitoring of blood-glucose: method for improving diabetic control. The Lancet, 1978, 311(8067): 729-732.
- [49] GUILLEN S, ARREDONDO M T. CASTELLANO E. A Survey of Commercial Wearable Systems for Sport Application [D]. In Wearable Monitoring Systems (pp. 165-178). Springer US, 2011.
- [50] SCHAFFER, CULLEN. Selecting a classification method by cross-validation. Machine Learning, 1993, 13(1): 135-143.
- [51] FRANTZIDIS, CHRISTOS A, CHARALAMPOS B. On the classification of emotional biosignals evoked while viewing affective pictures: an integrated data-mining-based approach for healthcare applications. IEEE Transactions on Information Technology in Biomedicine, 2010, 14(2): 309-318.
- [52] KAUR H, WASAN S K. Empirical Study on Applications of Data Mining Techniques in Healthcare [J]. Journal of Computer Science, 2006, 2: 194-200.
- [53] MOORHEAD, SUE A, JOANNE C. Nursing Interventions Classification: A Comparison with the Omaha System and the Home Healthcare Classification. Journal of Nursing Administration, 1993, 23(10): 23-29.
- [54] GROND, MARTIN, MAREK J. Improved Detection of Silent Atrial Fibrillation Using 72-Hour Holter ECG in Patients With Ischemic Stroke A Prospective Multicenter Cohort Study. Stroke, 2013, 44(12): 3357-3364.
- [55] JOHANNESSEN, LARS, ULRIC S L. A wavelet-based algorithm for delineation and classification of wave patterns in continuous Holter ECG recordings. In Computing in

- Cardiology, 2010 (pp. 979-982). IEEE.
- [56] MOSTELLER, FREDERICK, JOHN W T. Data analysis and regression: a second course in statistics. Addison-Wesley Series in Behavioral Science: Quantitative Methods. 1977.
- [57] 樊震林,黎爱军,吴宏.医疗风险影响因素的有序多分类 Logistic回归分析[J].中国卫生质量管理,2009, 16(4): 11-13.
- [58] KING, MATTHEW W, RESICK P A. Data Mining in Psychological Treatment Research: A Primer on Classification and Regression Trees. Journal of Consulting and Clinical Psychology, 2014.
- [59] HANRAHAN, KIRSTEN. From Research Results to Prediction and Translation: A Decision Support System for Children, Parents, and Distraction During Healthcare Procedures. In Sigma Theta Tau International's 23rd International Nursing Research Congress. STTI. 2012.
- [60] ANDERBERG, MICHAEL R. Cluster analysis for applications (No. OAS-TR-73-9). OFFICE OF THE ASSISTANT FOR STUDY SUPPORT KIRTLAND AFB N MEX. 1973.
- [61] ANDREA, LIVIA M, GARY L. Cluster analysis of adult children of alcoholics. Substance Use & Misuse, 1994, 29(5): 565-582
- [62] HASTIE, BARBARA A, JOSEPH L. Cluster analysis of multiple experimental pain modalities. Pain, 2005, 116(3): 227-237.
- [63] CLATWORTHY, JANE, DEANNA B. The use and reporting of cluster analysis in health psychology: A review. British journal of health psychology, 2005, 10(3): 329-358.
- [64] BOTSIS, TAXIARCHIS, MICHAEL D. Text mining for the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection. Journal of the American Medical Informatics Association, amiajnl-2010.
- [65] BORTOLAN G, DEGANI R, WILLEMS J L.. ECG classification with neural networks and cluster analysis. In Computers in Cardiology 1991, Proceedings. (pp. 177-180). IEEE.
- [66] YEH, YUN C, CHE W. Analyzing ECG for cardiac arrhythmia using cluster analysis. Expert Systems with Applications, 2012, 39(1): 1000-1010.
- [67] JUNG, SUN J, CHANG S. Association Rules to Identify Complications of Cerebral Infarction in Patients with Atrial Fibrillation. Healthcare informatics research, 2013, 19(1): 25-32.
- [68] HUANG, YI C. Mining association rules between abnormal health examination results and outpatient medical records. Health Information Management Journal, 2013, 42(2): 23.
- [69] DUAN, LIAN, NICK W. Healthcare information systems: data mining methods in the creation of a clinical recommender system. Enterprise Information Systems, 2011, 5(2): 169-181.
- [70] KHALILIA, MOHAMMED, SOUNAK C. Predicting disease risks from highly imbalanced data using random forest. BMC medical informatics and decision making, 2011, 11(1): 51.
- [71] MAZOMENOS, EVANGELOS B, DWAIPAYAN B. A low-complexity ECG feature extraction algorithm for mobile healthcare applications. IEEE journal of biomedical and health informatics, 2013, 17(2): 459-469.
- [72] PATIL, DIPTI D, SHAMLA T M. Feature Extraction Techniques For Mining ECG Signals In WBAN For Healthcare Applications. International Journal of Advances in Computing and Information Researches, 2012, 1(1): 11-15.
- [73] VERMA, NAVEEN, ALI S. A micro-power EEG acquisition SoC with integrated feature extraction processor for a chronic seizure detection system. IEEE Journal of Solid-State Circuits, 2010,45(4): 804-816.
- [74] DOULAH, ABUL B M, SHAIKH A F. DCT domain feature extraction scheme based on motor unit action potential of EMG signal for neuromuscular disease classification. Healthcare Technology Letters, 2014, 1(1): 26-31.

- [75] AGGARWAL, NAMITA, BHARTI R. Classification of alzheimer' s from t2 trans-axial brain mr images: A comparative study of feature extraction techniques. International Journal of Computer Vision and Image Processing (IJCVIP), 2012, 2(3): 30-43.
- [76] BANAEI, HADI, MOBYEN U A. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. Sensors, 2013, 13(12): 17472-17500.
- [77] KOH, HIAN C, GERALD T. Data mining applications in healthcare. Journal of Healthcare Information Management, 2011, 19(2): 65.
- [78] LEE, LIAO, EMBRECHTS M. Data mining techniques applied to medical information. Informatics f
- [79] BRESSAN, NADJA, ANDREW J. Integration of drug dosing data with physiological data streams using a cloud computing paradigm. In Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE (pp. 4175-4178). IEEE.
- [80] KAI, EIKO, ASHIR A. Technical challenges in providing remote health consultancy services for the unreached community. In Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on (pp. 1016-1020). IEEE.
- [81] STEVEN E, DILSIZIAN, ELIOT. Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. Current Cardiology Reports, 2013, 16:441
- [82] SHOUVAL R, BONDI O, MISHAN H.. Application of machine learning algorithms for clinical predictive modelling: a data-mining approach in SCT. Bone marrow transplantation, 2013, 49(3): 332-337.
- [83] GREEN K, STILPHEN ,.VILENSKY S. Computerized Clinical Decision Support System For Early Identification Of Patients Appropriate For Rehabilitation Services Improves Functional Status In Survivors Of Critical Illness. Am J Respir Crit Care Med, 2013, 187: A3621.
- [84] GULTEPE E, GREEN J P, NGUYEN H.. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. Journal of the American Medical Informatics Association, 2014, 21(2), 315-325.
- [85] KAFALI O, BROMURI S, SINDLAR M. Commodity 12: A smart e-health environment for diabetes management. Journal of Ambient Intelligence and Smart Environments, 2013, 5(5): 479-502.
- [86] MONEO G, DANIEL. Mobile devices in applications for healthcare: systems and technology.
- [87] VARNOSAFADERANI S R. The impact of ultra-fast broadband on telehealth in New Zealand: a thesis presented in partial fulfilment of the requirements for the degree of Master of Information Sciences in Information Technology at Massey University, Albany, New Zealand.(Doctoral dissertation), 2013.
- [88] ALGHAMDI R A. Telehealth practice and the impact of new technologies (NTs): a thesis submitted to the Institute of Information and Mathematics Sciences (IIMS) at Massey University in Albany campus in partial fulfilment of the requirements for the degree of Masters of Information Technology, Albany, Auckland, New Zealand, 2012.
- [89] TRIFIRO G, PARIENTE A, COLOMA P M. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? [J]. Pharmacoepidemiology and drug safety, 2009, 18(12): 1176-1184.
- [90] KAO, ROWLAND R, DANIEL T. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends in microbiology, 2014, 22(5): 282-291.
- [91] BROSSETTE, STEPHEN E, ALAN P.. Association rules and data mining in hospital infection control and public health surveillance. Journal of the American medical informatics association, 1998, 5(4): 373-381.
- [92] Wu, J., Roy, J., & Stewart, W. F. (2010). Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Medical care,

- 48(6), S106-S113.
- [93] JENSEN, PETER B, LARS J. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, 2012, 13(6): 395-405.
- [94] ENGEL, GEORGE L. The need for a new medical model: a challenge for biomedicine. *Science*, 1977, 196(4286): 129-136.
- [95] EVANS, STEVEN, STEPHEN J. Automated detection of hereditary syndromes using data mining. *Computers and biomedical research*, 1997, 30(5): 337-348.
- [96] HUANG, QI R, ZHENXING Q. Clinical patterns of obstructive sleep apnea and its comorbid conditions: a data mining approach. *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 2008, 4(6): 543.
- [97] ZRIMEC, TATJANA, JAMES. Improving computer aided disease detection using knowledge of disease appearance. In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems* (p. 1324). IOS Press, 2007.
- [98] MELZER, TRACY R, RICHARD W. Arterial spin labelling reveals an abnormal cerebral perfusion pattern in Parkinson' s disease. *Brain*, awq377, 2011.
- [99] YUANGYUANG X, QI L, LI J. Detecting adolescent psychological pressures form Micro-Blog. *Health Information Science*, 2014, LNCS8423: 83-94.
- 收稿日期: 2014年6月30日
- 颜 延: 中国科学院深圳先进技术研究院, 博士研究生, 主要研究方向为机器学习、数据挖掘与生物医学信息处理。
E-mail: yan.yan@siat.ac.cn
- 秦兴彬: 中国科学院深圳先进技术研究院, 硕士研究生, 主要研究方向为计算机应用技术。
E-mail: xb.qin@siat.ac.cn
- 樊建平: 中国科学院深圳先进技术研究院, 研究员, 博士生导师, 主要研究方向为高性能计算及计算机应用技术。
E-mail: jp.fan@siat.ac.cn
- 王 磊: 中国科学院深圳先进技术研究院, 研究员, 博士生导师, 主要研究方向为人体传感器网络 (Body Sensor Network, BSN) 及其在低成本健康和手术机器人中的集成应用。
E-mail: wang.lei@siat.ac.cn