医学大数据应用前景与挑战

王雪梅 刘 莉

(中南大学信息安全与大数据研究院医药信息系 长沙 410083)

[摘要] 阐述医学大数据的特征, 分析医学大数据在临床预测、流行病学监测、临床检验、临床诊疗等方面的应用前景, 指出政府部门、患者、医护人员及健康人群等受益群体及面临的技术标准、数据管理等方面的挑战, 同时提出充分挖掘和利用医学大数据的重要性。

[关键词] 大数据; 医学; 应用前景; 挑战

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2016.08.013

Application Prospect and Challenges of Medical Big Data WANG Xue - mei, LIU Li, Department of Medical Information, Institute of Information Security and Big Data, Central South University, Changsha 410083, China

[Abstract] The paper elaborates on characteristics of medical big data, analyzes the application prospect of medical big data in clinical prediction, epidemiological monitoring, clinical examination and clinical diagnosis and treatment, and points out the beneficiary groups such as governmental departments, patients, medical personnel and healthy people and the potential challenges. Also, it indicates the importance of fully excavating and utilizing medical big data.

[Keywords] Big data; Medical; Application prospect; Challenge

1 引言

大数据是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合, 是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化功能来适应海量、高增长率、多样化的信息资产。近年来, 随着大数据热潮的兴起, 大数据已渗透到每一个行业和业务领域, 成为重要的生产要素。在医疗领域, 早期的数据是以纸张化的形式存在的, 随着强大的数据存储、计算平台及移

动互联网的发展, 医学数据呈现大量爆发及数字化的趋势; 同时数据的来源也纷繁复杂, 可能来自不同的地区、不同的医疗机构、不同的软件应用。鉴于此本文分析了大数据在医学领域的应用前景与挑战。面对海量杂乱无章、多格式、多源头、爆炸性增长的医疗大数据, 一旦能对其进行科学的整理和分析, 将对提高医疗质量、增强患者安全、降低风险及医疗成本等方面产生无以伦比的巨大的作用。

2 医学大数据的特征

关于大数据的特征, 业内专业人士表示可以用很多词语表达, 其中比较有代表性的即为 IBM 公司提出的“5V”特性, 即大量化 (Volume)、多样化 (Variety)、价值化 (Value)、快速化 (Velocity) 和

[收稿日期] 2016-04-09

[作者简介] 王雪梅, 在读硕士研究生; 刘莉, 副教授, 发表论文多篇。

真实化 (Veracity)^[1]。在对大数据的研究与应用中,掌握这 5 个特性即可充分利用大数据的价值。2015 年 8 月 31 日,国务院在印发的《促进大数据发展行动纲要》^[2]中指出:“坚持创新驱动发展,加快大数据部署,深化大数据应用,已成为稳增长、促改革、调结构、惠民生和推动政府治理能力现代化的内在需要和必然选择”。因此,深化大数据应用,充分发挥大数据的价值显得尤为重要。医学大数据是海量数据中的一个分支,不仅具有上述“5V”特性,而且还有其特殊性,具体体现在:(1) 医学数据的增长速度更快、更复杂,数据的半结构化、非结构化特征更加明显;尤其是近年来随着各种健身、健康可穿戴设备等的出现,使得血压、心率、体重、血糖、心电图 (EKG) 等指标的实时监测成为可能,信息获取和分析的速度已经从原来的按“天”计算发展到了按“小时”、“秒”计算。(2) 医学数据要求更加准确,疾病诊断和治疗不存在模棱两可的说法,如肺炎患者不能被诊断为可能肺炎或可能流行性感冒,因为就医和诊治强调的是医生能够对症下药。(3) 医学数据更强调数据的时效性,如对于急救疾病必须分秒必争、对于死亡记录时间必须足够精确等。

3 医学大数据应用前景

3.1 临床预测

通过大量临床数据分析可以为临床决策提供支持。如基于大量结构化、非结构化临床数据分析,医生通过比对已有结果与现有患者的相关数据、指标,即可对患者的疾病状况、住院时间及手术最佳时间等进行估计,患者只要完善较少的相关检验、检查项目即可获得高效、精确的治疗。同时,通过已有数据的分析,对临床现有的疾病预测模型及方法、并发症评估模型等的设计和开发提出了更高的要求,要建立成本更低、效率更高、更精简、更快速、针对性更强的模型,以更准确地预测疾病发生率、并发症发生率等。

3.2 流行病学监测

基于大量流行病学数据的病种分析,预测流行

性疾病爆发对于控制其蔓延至关重要,尤其是对于那些既没有疫苗预防、也没有有效方法治疗的疫情更加有效。如在登革热流行期间,哈佛大学 Wesolowski 等通过整合 2013 年大约 4 000 万巴基斯坦手机用户 (占全国人口的 22%) 的天气和移动数据,预测出了登革热的蔓延趋势^[2]。另外,利用疾病地图的相关软件和分析方法,对特定病种的大量流行病数据进行分析,对疫情监测、疾病的地区差异、时间差异等进行评估,以预测疾病发生,对改善人们的生活方式和生活质量至关重要。

3.3 临床检验

以大量临床检验、检查数据为基础,对病种的检验、检查项目进行分类整理,综合分析比较各种疾病检验、检查项目的必要性,通过数据以及临床医生经验的验证,以减少现有的所谓必要的检验、检查项目,进而降低患者医疗费用,减少检验科的患者量,减轻医生的负担,同时降低医疗资源消耗。另外,以影像存储与传输等系统为基础,以相关国际标准和医学代码为依据,对疾病检验、检查项目进行结构化,进而适应大数据环境下医学数据的互联互通,发挥检验、检查项目隐含的价值。

3.4 临床诊疗

通过对疾病的研究与预测,结合相关必要项目的检查,能够为医生在疾病诊断方面提供参考。通过对大量现有临床诊疗数据进行分析,比较同病种不同治疗措施的有效性,医生能够选择更优的与病症信息相匹配的治疗方案。且通过多类数据 (临床治疗、基因组测序、组织形态等) 分析,医生可以为每位患者量身定制治疗方案,使个性化医疗成为可能。

3.5 药品和医疗器械研发

通过对大量医疗器械、临床数据、流行病学资料等医学数据的分析,可以对现有的药品和医疗器械研发进行改进,研制出更低磨损度、更精简、更快速、更有针对性、更低成本的产品,以提高诊断和治疗的准确性。另外,根据分析和预测,研究人

员可以更加准确地研制出疾病的新靶向疫苗、新分子标记物等新技术、新药品,降低患病率,提高生存率。

3.6 医院管理

行政管理人员通过对医院现有数据的分析,可以为医院决策、管理提供支持。目前北京公立医院管理层面针对大数据应用做了很多工作,统一规划、统一标准、统一实施。医学大数据的充分利用不仅能够加强医院的管理,促进医院改革向现代化管理制度发展,而且还能控制医院的收益,保证收支平衡;而这个层面上也对医疗软件企业管理或财务平台等的构建提出了更高的要求。医院行政管理部门根据大数据分析结果,可以更好地安排全院工作,发挥其监督、管理等职能。

3.7 生物医学

目前医学大数据较多地应用在生物信息学方面,研究者认为可以通过医学大数据、普通人群健康信息等分析,达到更有效和低成本地执行基因组、转录组测序,使基因组分析成为正规医疗保健决策的必要信息并纳入病人病历记录;但是由于生物信息处理需要复杂的信息提取工具和算法、生命系统本身极其复杂,加之数据获取困难和高昂的代价^[3],导致生物医学大数据的充分利用存在巨大挑战。

3.8 健康诈骗

卫生防疫、卫生监督等机构通过对人群健康大数据的分析,可提前预测公民健康状况,用于市场监管,以防止公民被骗而采购不必要的营养保健品,进而减少消费者的经济和精神负担。另外,保险公司可以通过快速分析大量索赔请求,结合健康大数据分析结果,判断索赔理由的真实性,降低欺诈成功率,减少浪费和滥用^[4]。

3.9 远程监控

对于疑似患病者或慢性病患者等,可以通过采集和分析患者医院或家庭实时大容量快速移动数

据,用于安全监控和不良反应预测。如自中南大学临床大数据项目开展以来,中南大学已研制出流式健康测量和健康信息推送两套大数据原型系统,前者可在手机、平板电脑等智能终端上对脉搏、血压、心电等人体生理指标进行测量,将结果传送到湘雅大数据中心;后者可让患者从湘雅大数据中心获取自身的各项生理指标测量结果并及时推送到智能终端上,使患者获得更便捷的医疗信息服务^[5]。

4 医学大数据受益群体

4.1 政府部门

卫生计生相关部门以医学大数据分析结果为基础,能够更加方便、快捷地管理传染病等流行病以及公共卫生等问题。如大量传染病数据的分析,可以及早发现新发传染病,掌握人流动向进而切断传播途径,及早进行疫苗的准备,最终达到全民及早参与的局面;通过将环境大数据和人群健康大数据相结合,可以快速分析环境对健康的影响,以及环境污染与某些疾病的关联性;通过医学大数据分析,相关部门可以进行公共卫生舆情监控预警,对医疗资源进行适当分配,避免不合理分配造成的医疗资源不足或浪费;通过寻医问药数据、患者诊疗数据以及健康数据的分析,可以得出食品安全问题与疾病的关联性。

4.2 患者、医护人员以及健康人群

对于患者来说,通过“前辈”对众多经验的分享,可以指导其便捷就医;通过医生对已有数据分析的结果,可以获得个性化的医疗服务;对于手术患者或特殊疾病患者,通过相关疾病预测模型分析,可以预测相关患者并发风险、手术最佳时期及最佳方式、预后情况,以提高治愈率、降低风险;对于出院病人、慢性病患者、就医不便者和疑难杂症者可以使远程医疗监控成为可能。这些都在一定程度上减轻了患者及其家属的经济和精神负担。对于医护人员来说,通过疾病分析与预测,护理人员可以提早进行干预并可为临床医生提供辅助诊断,提高诊疗效果。对于健康人群来说,通过对大量健

康数据的分析,可以对患者选择加强锻炼还是入院诊治,或是保健药等提供依据,从而提高人们的生活质量。

4.3 医疗机构

大数据环境下,在强有力的技术、政策、专家团队等支持下,使得医院之间的信息能够共享,达到真正的互联互通,有效减少各医疗机构之间医疗信息不对称的现象。通过分析大中小型医院现有的医疗大数据,为各医院诊疗提供决策支持,为二级以下医院甚至社区医院或诊所疑难疾病的诊疗提供依据。另外,根据医院各部门的大数据分析结果,管理部门可以控制医院成本,保证医院收支平衡,且可以充分调动医院资源,进行人员、设备等的合理分配;药剂科、设备科可以进行药物、设备选购数量、品牌等的上报及选购;二级以上已实施 DRGs 医院的收费科室可以上报保险公司,保险部门可以对相应疾病的医保措施进行改进等。

4.4 科研机构

科研机构不仅包括医院内临床病种相关的科研小组、基础医学相关科研机构,还包括第 3 方科研机构,各科研机构从医疗大数据分析中获得的益处是显而易见的。如目前医院医生职称评选,既要求进行临床,又要有一定的科研成果;而有了医学大数据平台,医生不再需要像以前一样顶着巨大压力进行科研,只要充分利用该平台中的数据进行分析即可。医学研究中很常用的大样本随机对照试验(RCT)对试验本身有严格的要求,主观性的特点使试验结果存在巨大偏差;但是通过对医学大数据中多种病种数据的分析,可以打破传统的经验医学和大样本随机对照临床试验分析,克服 RCT 的主观性和结果差异,为科学研究提供支撑,进而为临床决策提供支持。

4.5 公司

医药企业可以根据临床、药品、器械等大数据分析当前生产的药物是否符合市场需求,帮助药企对药品生产及淘汰提供支持;也可以根据未来大数

据环境下部分信息公开情况,研制高品质、高效率的产品,提高医疗质量。保险公司根据健康大数据分析结果可以提前检测和防止健康诈骗,也可以为投保人设计个性化保险。医疗器械公司和穿戴设备公司在自身产品设计与分配上也能从中得到一定的启发和利益。

5 医学大数据面临的挑战

5.1 大数据“自大”

大数据所谓的“大”并非数据量越多越好,而是要力求数据多中求优,甄选对分析有价值的数;尤其对于病历数据中众多的信息内容,在分析之前应该拟定好要分析的项目,进而获取有用的信息。另外在数据分析的过程中,由于样本的特殊性,仍存在过度拟合的问题,最佳预测模型并非适用于所有的研究样本,这对研究者提出了更高的要求。

5.2 因果关系被忽略

在大数据分析中,研究人员主要关注的是数据之间的相关关系,却忽略了数据之间最重要的因果关系。在研究和应用中人们应该更加注重大数据中的数据生成机制,对数据生成机制相同的数据进行分析,其预测结果更具准确性、代表性。如在临床上不能应用糖尿病大数据分析得到的糖尿病并发症预测模型来预测冠心病并发症的发生,虽然其中部分指标会存在交叉,但是二者的发病机制存在差异,部分特异性指标也就不同。另外,对于数据收集者和数据使用者,应该加强与数据生产者与记录者之间的沟通,如研究人员、数据分析人员应该与临床医生加强沟通,利用数据挖掘技术采集有用指标。

5.3 技术与标准

传统的数据挖掘技术、计算机统计技术以及相关统计方法,已不能满足大数据时代的要求,在对大数据进行分析的过程中对研究人员、技术人员提出了更高的要求。不同地区、不同医生使用的医学术语可能存在差别,医生未使用结构化语言,对于海量数据的采集处理很难达到数据的利用与共享。

所以使用标准化的结构化语言对于医学大数据采集尤为重要。

5.4 数据管理

在大数据环境下,呈爆发式增长的数据在目前的管理机制下无法达到最优化管理,需要存储量更大的电子介质、更专业的管理人员、更先进的数据处理技术等对其进行专业、高效管理,代价过高。企业所构建的医院管理平台具有专有性,只是针对部分或单个医院,医院管理未达到标准化、统一化。

6 结语

充分发掘和利用医学大数据的价值至关重要,不仅能为临床医生、医保机构提供决策支持,为医院管理提供依据,而且还能为患者解决看病难、看病贵的问题,为家庭减轻经济负担,进一步为社会

减轻经济压力。但是同时也要认识到对医学大数据的分析还面临诸多挑战,尤其是数据“自大”、技术跟不上、管理困难等问题,对医学大数据的分析与利用任重而道远。

参考文献

- 1 李清泉,李德仁. 大数据 GIS [J]. 武汉大学学报(信息科学版), 2014, 39(6): 641-644.
- 2 Wesolowski A, Qureshi T, Boni MF, et al. Impact of Human Mobility on the Emergence of Dengue Epidemics in Pakistan [J]. Proc Natl Acad Sci USA, 2015, 112(38): 11887-11892.
- 3 宁康,陈挺. 生物医学大数据的现状与展望 [J]. 科学通报, 2015, 60(5/6): 534-546.
- 4 陈遵秋. 浅谈医学大数据 [EB/OL]. [2015-03-23]. http://www.china-cloud.com/yunji-shu/shujuzhongxin/20150323_45875.html.
- 5 张惠. 中南大学建临床医学大数据项目 [N]. 中国信息化周报, 2014-2-10(001).

(上接第 52 页)

- 4 黄逸民. 基于多 Agent 的智能管理信息系统理论与应用研究 [D]. 杭州:浙江大学, 2002.
- 5 余腊生,李礼. 基于 Agent 的医疗信息整合平台的设计与研究 [J]. 计算机工程与设计, 2010, 31(19): 4197-4200.
- 6 张春飞,李万龙,许志军,等. 基于 Agent 的医疗信息整合系统模型 [J]. 河北工业科技, 2009, 26(6): 484-488, 496.
- 7 Chenglang Lu, Zongda Wu, Mingyong Liu, et al. A Patient Privacy Protection Scheme for Medical Information System [J]. Journal of Medical Systems, 2013, (37): 9982-9991.
- 8 赵晓佩. 电子病历与患者隐私权保护探究 [J]. 质量与信息化, 2011, 18(5): 61-63.

- 9 乔诚,王碧琴,张宏静. 基于信息共享系统的信息隐私技术保护体系 [J]. 兰台世界, 2012, (11): 54-55.
- 10 赵君. 医院信息化之 LIS 与 HIS 无缝连接 [J]. 中国病案, 2009, 10(10): 32-33.
- 11 孙波,刘国强. 信息化医院 HIS 系统与 LIS 系统的集成研究与实现 [J]. 医院管理, 2009, 4(21): 113-117.
- 12 陆骥. LIS 与 HIS 的系统集成设计 [J]. 医院数字化, 2005, 26(2): 36-37.
- 13 吴寿刚,王晓华. 基于 HL7 的医疗信息系统集成应用 [J]. 中华医学图书情报杂志, 2014, 23(1): 60-64.
- 14 关延风,马骋宇. 基于电子病历的医疗信息隐私保护研究 [J]. 医学信息学杂志, 2011, 32(8): 36-39.
- 15 刘甘林. 基于 Java 的软件 Agent 实现技术研究 [D]. 长沙:国防科学技术大学, 2004.