

急救大数据与 Datathon 活动

张政波 薛万国 曹德森 黎檀实

100853 北京,解放军总医院医学工程保障中心(张政波、曹德森),医疗大数据中心(张政波、薛万国),急诊科(黎檀实)

通讯作者:黎檀实, Email: lts301@163.com

DOI: 10.3760/cma.j.issn.2095-4352.2018.06.020

【摘要】 急诊危重症患者在诊疗过程中会产生大量、高密度的临床数据,这些数据的整合和分析利用,对于提高救治水平及效率,发展基于数据驱动的临床辅助决策支持系统具有重要价值。我国虽然拥有丰富的病例资源,但在临床数据库建设及其分析利用等方面才刚刚起步。解放军总医院在急诊数据库建设及数据共享利用方面进行了积极的探索,建立了国内首个急救数据库,并基于该数据库举办了跨学科合作的 Datathon 数据实践活动,为国内临床数据整合、数据库建设、跨学科协作及数据共享提供了借鉴经验。本文结合国内外进展情况,对该项工作进行总结评述,并进一步提出建立急救医疗大数据联盟及构建学习型医疗系统等倡议,以汇聚更多的临床数据资源,并将临床数据库建设-分析-应用形成闭环,使医疗大数据在降低医疗费用和提高医疗质量等方面切实发挥作用。

【关键词】 医疗大数据; 急救; 跨学科合作; Datathon; 学习型医疗系统

基金项目:国家自然科学基金(61471398);北京市科委医药协同科技创新研究项目(Z181100001918023);解放军总医院医疗大数据研发项目(2016MBD-027);医疗大数据应用技术国家工程实验室(2017-148)

Big data in emergency medicine and Datathon event Zhang Zhengbo, Xue Wanguo, Cao Desen, Li Tanshi
Department of Biomedical Engineering and Maintenance Center, Chinese PLA General Hospital, Beijing 100853, China (Zhang ZB, Cao DS); Medical Information Center, Chinese PLA General Hospital, Beijing 100853, China (Zhang ZB, Xue WG); Department of Emergency, Chinese PLA General Hospital, Beijing 100853, China (Li TS)
Corresponding author: Li Tanshi, Email: lts301@163.com

【Abstract】 A detailed, high-scale clinical data can be generated in the process of diagnosis and treatment of emergency critically ill patients. The integration and analysis and utilization of these data are of great value for improving the treatment level and efficiency and developing the data-driven clinical assistant decision support. China has large volume of health information resources, however, the construction of healthcare databases and subsequent secondary analysis has just started. With the effort of the Chinese PLA General Hospital in building an emergency database and promoting data sharing, the first emergency database was published in China and a health Datathon was organized utilizing this database, providing experience for clinical data integration, database construction, cross-disciplinary collaboration and data sharing. Referring to the development at home and abroad, this review discussed work in this area and further proposed establishing a big data cooperation for emergency medicine and building a learning healthcare system to integrate more clinical resources and form a closed loop of "clinical database construction-analysis-applications", and enhance the effectiveness of medical big data in reducing medical costs and improving healthcare delivery.

【Key words】 Big data in healthcare; Emergency medicine; Cross-disciplinary collaboration; Datathon; Learning healthcare system

Fund program: National Natural Science Foundation of China (61471398); Beijing Municipal Science and Technology Project (Z181100001918023); Big-data Research and Development Project of Chinese PLA General Hospital (2016MBD-027); National Engineering Laboratory for Industrial Big-data Application Technology (2017-148)

当前医疗技术日新月异,随着数字技术的发展,临床环境中产生了海量的、不同形式的医疗数据,我们也进入了“医疗大数据时代”。哈佛医学院、斯坦福医学院都相继提出了“大数据将改变未来的医疗模式”“培养下一代懂得医疗大数据的医护人员”“需要进一步培养医生和患者的数据管理及分析技能”,让更多的临床医护人员了解医疗大数据,进而在医疗实践中使用医疗大数据。我国虽然拥有丰富的病例资源,但在临床数据库建设及其分析利用等方面才刚刚

起步,如何整合多源异构临床数据资源,形成高质量临床数据库,进而在此基础上进行二次分析利用,为临床精准医疗、科学研究、运营管理提供辅助决策支持,亟需新的思维、方法和工具。

解放军总医院急救数据库建设及第一届急救大数据 Datathon 活动,在临床数据库建设及共享利用方面进行了大胆探索和尝试,为国内临床数据资源整合和利用提供了很好的借鉴经验。现结合国内外进展情况,对该项工作进行评述。

1 急救数据库建设

急救医学始终冲锋在抢救生命的最前沿,急诊患者疾病发展迅速,病情变化快,因此对疾病实现早识别、早诊断、早救治是急危重症救治的关键,在急救过程中产生大量高密度的临床数据,整合和分析利用这些数据对提高救治水平,发展基于数据驱动的临床辅助决策支持系统具有重要价值。由于急诊中有大量创伤患者,整合和利用这些患者的数据,对于发展战创伤救治技术也具有重要价值。

急诊可以看作是一个涵盖了多专科的复杂小型医院,患者疾病谱广,业务流程比普通临床科室复杂,因此,急救数据库的建设具有一定的挑战性。急危重症数据库数据主要来源于解放军总医院急诊部门、专科诊室和医院信息系统(HIS)。急诊部门和专科诊室数据分别属于不同的医疗系统,从急诊信息系统获得患者基本信息、分诊信息、就诊区域及其床旁监护数据;从专科系统获取患者就诊过程中的医嘱、用药和护理记录等信息;从HIS获得患者的化验检查数据。根据急诊患者的就诊流程,将患者从挂号分诊到院内急救和转出这一过程中产生的一系列数据进行有效关联,实现数据整合。该数据库的数据表结构设计参考了重症监护医学信息数据库(MIMIC)^[1],可分为临床业务表、临床信息表和字典表三大类,每一类又有若干表,该工作也得到了麻省理工学院和哈佛医学院相关人员的技术支持。急救数据库涵盖了解放军总医院从2015至2018年抢救单元收治的急危重症患者诊疗数据,包含50万例次患者分诊信息以及近3万例次抢救间救治数据,包括详细的诊断、治疗、生理、生化、影像、医嘱和护理记录等信息,为该领域的相关研究提供了来自“真实世界”的宝贵数据。

数据的价值在于利用,解放军总医院急救数据库建设为国内同行提供了参考,后续会进一步推动该领域的相关工作,实现不同医疗机构之间的数据资源建设和共享利用,建立急救医疗大数据联盟。未来急救大数据的发展,是整合更多医疗机构的急诊数据资源,并整合院前、院内、院外随访数据,形成多中心全信息链急救大数据。发展医学人工智能的三大要素为算力、算法和大数据,相信多中心全信息链急救大数据的建设将进一步助力医学人工智能的发展,将有越来越多的人工智能技术应用于疾病的预测、预警和辅助决策支持。

2 跨学科协作的 Datathon 活动

Datathon=Data+Hackthon。Datathon的前身源于硅谷的“Hackathon”。Hackathon通常是以短期、高强度的小组竞赛形式进行,旨在促进各大公司的技术创新和新观念的产生。对于医疗数据分析来说,Hackathon的目的在于汇集知识背景和技术不同但互补的参与者,包括临床专家、数据科学家、统计学家、工程师和计算生物学家等,通过使用来自“真实世界”的医疗数据,合作解决临床医生面临的众多问题和未满足的医疗需求,并产生可靠的、可复制的研究结果。基于这个目的,“Data”和“Hackathon”结合形成Datathon^[2]。

典型的Datathon活动一般为期2d(周六、日),所有参加人员根据个人兴趣和专业背景分为不同的小组,每个小组集中解决一个基于医疗大数据分析的临床研究问题。Datathon开始前通常是“路演”环节,Datathon队长上台介绍研究题目和内容,招募队员,参赛人员在现场根据研究题目和自己的专长进行选题及组队。每个Datathon小组必须有2~3名临床医生、3~5名数据工程师和数据科学家,其中研究问题的提出多来自临床医生,数据工程师负责数据提取、清洗、可视化等工作,数据科学家负责建模和统计分析,通过不同专业背景人员之间的思维碰撞和分工协作,探索基于医疗数据来回答临床研究问题。在这个过程中可以学习和体会如何定义临床问题以及提取数据的相关技巧。在“Hacking”环节,参赛人员被鼓励去分享他们的数据提取和分析代码(如MATLAB、Python、R、SAS等),其中包括数据提取、统计分析、机器学习算法模型等。Datathon的最后环节是成果展现,参赛人员介绍他们的项目进展并且聚焦于基于“真实世界”数据开展研究的挑战以及这次活动的心得体会,评委会给出评审意见并颁奖。

Datathon活动最初由麻省理工学院和哈佛医学院的团队发起,麻省理工学院计算生理学实验的团队在2014年和2015年分别举行了2次国际性的Datathon活动。迄今为止,该活动已在全球举办了近20场,其中在中国举办了3场,由解放军总医院团队最先将其引入国内^[3]。Datathon是一项新颖的数据实践活动,是促进医疗大数据落地应用的有效形式,它既是一项竞赛,也是一种培训。通过Datathon过程,临床人员基于临床经验和需求提出问题,将临床问题转化为研究问题(即确定研究队列、暴露变量和结局变量等),进而将其转化为工程人员能够理解的语言,使工程人员能够有效地从数据库中提取目标数据;工程人员需准确地解读临床需求,确保临床需求是清晰明确的,并且能够转化为具体的数据操作,之后操作数据库,提取数据并进行分析处理,建模或可视化,并将其结果反馈给临床人员。在整个Datathon活动过程中,临床人员与工程人员的这种互动是持续的,伴随着数据提取和分析的每一步,可能都需要二者一起探索、头脑风暴、寻找解决问题的办法。在这个过程中,还会有经验丰富的场外人员(临床专家和数据科学家)参与讨论,给予辅导。

Datathon活动相当于将一项基于医疗大数据的研究过程浓缩到2d,通过为期2d的数据实践活动,参与队伍基本上都能完成一个“定义研究问题—提取数据—分析建模—结果验证”的流程,快速学习和掌握医疗大数据分析的基本思路、方法及技能,体验不同学科背景人员的分工协作、思维碰撞和成果分享过程,体会到为什么医疗大数据应用需要多学科合作以及如何扮演好各自的角色。

2018年3月举办的第一届急救大数据Datathon活动是国内首个基于急救数据库的Datathon活动,也是国内首个使用自主临床数据库开展的Datathon活动,有13个团队、近

200人参加,竞赛题目包括:基于急救大数据的高乳酸血症短时长治疗效果预测模型建立研究;影像报告自动语意分析,呼吸衰竭患者结局相关因素分析;基于急救大数据的发热疾病早期诊断与病因预测;基于机器学习的急性冠状动脉综合征快速分诊评估;不同评分系统对急性重症胰腺炎预警作用比较,需紧急救治危重患者相关危险因素分析;基于多主体智能粗糙集筛选算法的消化道出血患者院内再出血治疗关键生理指标提取;急性感染患者炎症因子水平与病原学关联研究;消化道出血风险与气象因素关系研究;动脉血乳酸联合全身炎症反应综合征(SIRS)在急危重症患者早期评估预警中价值研究;基于非支配排序筛选算法的急危重症患者结局相关关键指标提取;基于Apriori算法的患者用药谱关联因素分析。此次活动意义深远,进一步推动了临床数据库的开放共享和落地应用。来自医疗机构的参赛人员纷纷表示,他们要联系其所在医院的IT部门和医工部门,启动自己医院的临床数据库建设工作;而工程人员则纷纷表示,希望后续能够与临床人员继续合作,基于急救数据库开展更多更深入的研究。

3 构建基于数据驱动的学习型医疗系统(LHS)

LHS由美国医学院提出,目的是衔接研究与临床应用,利用来自“真实世界”的数据产生临床证据,不断提高医疗系统的效能。医疗大数据分析利用的最终目的是建立基于数据驱动的LHS^[4-5]。

LHS以研究问题为起点,通常包含传入通道、传出通道和推广应用3个组成部分。传入通道收集医疗实践中产生的多源的数据,通过各种方法分析数据,并对发现作出解释,形成临床证据;传出通道则将上述发现以各种形式将结果反馈回医疗系统,改变医疗实践;推广应用则是考虑上述发现(临床证据)的有效性,无论是在医疗机构内部,还是全国乃至全球范围内都有效,具有可推广性。LHS意味着研究模

式的转变,即以患者的需要为研究目标。LHS要持续地收集和分析临床数据,并且将结果反馈到医疗实践中去,带来诊疗决策的持续优化和改进。相比传统的基础研究,LHS开展的研究更贴近临床需求,科学家、临床人员和管理人员通常协同工作,而且围绕着临床实践,研究工作每天都可进行。由于这些研究是基于“真实世界”的临床数据,因此,多源电子病历数据被整合和挖掘利用,并且大家都认识到临床和医疗数据对大众健康及医疗持续改进的重要价值,观察性研究和实用性随机对照临床试验(RCT)研究的应用范围将更为广阔,能够更快、更低成本地填补循证空白。

LHS的概念比医疗大数据更广,医疗大数据仅代表了LHS的传入通道,而LHS能够支持完整的学习闭环,将基于医疗大数据的发现反馈应用到医疗实践中去,不断地提高医疗质量。因此,临床数据库建设和Datathon活动仅是构建LHS的开始,后续还需多方力量协作,将临床数据库建设-分析-应用形成闭环,使医疗大数据在降低医疗费用和提高医疗质量方面切实发挥作用。

参考文献

- [1] Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database [J]. *Sci Data*, 2016, 3: 160035. DOI: 10.1038/sdata.2016.35.
- [2] Aboab J, Celi LA, Charlton P, et al. A "datathon" model to support cross-disciplinary collaboration [J]. *Sci Transl Med*, 2016, 8 (333): 333ps8. DOI: 10.1126/scitranslmed.aad9072.
- [3] Li P, Xie C, Pollard T, et al. Promoting secondary analysis of electronic medical records in China: summary of the PLAGH-MIT critical data conference and health datathon [J]. *JMIR Med Inform*, 2017, 5 (4): e43. DOI: 10.2196/medinform.7380.
- [4] Budrionis A, Bellika JG. The learning healthcare system: where are we now? A systematic review [J]. *J Biomed Inform*, 2016, 64: 87-92. DOI: 10.1016/j.jbi.2016.09.018.
- [5] Celi LA, Mark RG, Stone DJ, et al. "Big data" in the intensive care unit. Closing the data loop [J]. *Am J Respir Crit Care Med*, 2013, 187 (11): 1157-1160. DOI: 10.1164/rccm.201212-2311ED.

(收稿日期:2018-04-23)

• 科研新闻速递 •

冠状动脉内 CD34⁺ 干细胞移植治疗终末期冠心病患者的远期效果

有学者进行了一项临床研究,旨在了解冠状动脉(冠脉)内 CD34⁺ 干细胞移植治疗终末期冠心病患者的远期效果。研究人员将 38 例接受冠脉内 CD34⁺ 干细胞移植治疗的终末期冠心病患者随机分为两组:第一组患者($n=18$)每根血管内移植 10 个 CD34⁺ 干细胞,第二组患者($n=20$)每根血管内移植 30 个 CD34⁺ 干细胞。所有受试者均不适合接受介入治疗或冠脉旁路移植术。结果显示:受试者 5 年生存率和主要心脑血管事件发生率分别为 78.9% (30/38) 和 36.8% (14/38)。在随访期间,有 31.6% 的患者 (12/38) 在接受治疗后靶血管直径增大至可以接受介入治疗的标准,并最终接受了冠脉支架植入术。死亡患者血管内皮功能较存活患者显著下降 ($P=0.039$)。Wimasis 图像分析显示,受试者接受 CD34⁺ 干细胞移植治疗后血管增生明显。三维超声心动图显示,受试者左室射血分数在接受治疗后 1 年内持续升高,并在 5 年内保持稳定;而左心室的直径变化则表现出与左室射血分数相反的规律。患者心绞痛和心力衰竭(心衰)的临床评分持续下降。研究人员据此得出结论,冠脉内 CD34⁺ 干细胞移植治疗可持续改善终末期冠心病患者的左心功能,减少心绞痛和心衰发作,并可抑制左室重构。

罗红敏,编译自《*Crit Care Med*》,2018,46(5):e411-418