

推动医疗大数据应用落地的跨学科

Datathon 模式探索

张渊 李沛尧 赵宇卓 刘同波 张政波 曹德森 黎檀实 Leo Anthony Celi

100853 北京,解放军医学院(张渊);100853 北京,解放军总医院医学工程保障中心(李沛尧、张政波、曹德森),急诊科(赵宇卓、黎檀实),医疗大数据中心(刘同波、张政波),计算机应用与管理科(刘同波),医疗器械研发与评价中心(张政波、曹德森);02139 美国马萨诸塞州,麻省理工学院计算生理学实验室(Leo Anthony Celi);02138 美国马萨诸塞州,哈佛医学院 贝斯以色列女执事医学中心(Leo Anthony Celi)

通讯作者:张政波,Email:zhengbozhang@126.com

DOI: 10.3760/cma.j.issn.2095-4352.2018.06.021

【摘要】 医疗实践活动产生海量的数据,对这些数据的整合、分析和利用需要跨学科合作,需要新的模式和方法,以推动医疗大数据应用和人工智能研究的落地。Datathon 是一项新颖的数据实践活动,通过组建理-工-医结合的跨学科团队,以短期高强度的项目竞赛形式,基于“真实世界”的数据来探索解决临床研究问题。本文介绍了 Datathon 活动的背景起源、组织模式和相关实践。Datathon 活动为国内医疗领域的跨学科合作提供了创新协作的新模式,为如何开展医疗大数据研究提供了新思路,有利于不同学科间优势互补,凝聚多方力量一起探索解答临床实践中遇到的问题,并推动医疗大数据实践应用落地。

【关键词】 医疗大数据; 大数据应用; 跨学科合作; Datathon

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)

A cross-disciplinary collaborative "Datathon" model to promote the application of medical big data Zhang

Yuan, Li Peiyao, Zhao Yuzhuo, Liu Tongbo, Zhang Zhengbo, Cao Desen, Li Tanshi, Leo Anthony Celi

Chinese PLA Medical School, Beijing 100853, China (Zhang Y); Department of Biomedical Engineering and Maintenance Center, Chinese PLA General Hospital, Beijing 100853, China (Li PY, Zhang ZB, Cao DS); Department of Emergency, Chinese PLA General Hospital, Beijing 100853, China (Zhao YZ, Li TS); Medical Information Center, Chinese PLA General Hospital, Beijing 100853, China (Liu TB, Zhang ZB); Department of Computer Application and Management, Chinese PLA General Hospital, Beijing 100853, China (Liu TB); Department of Medical Device R&D and Evaluation Center, Chinese PLA General Hospital, Beijing 100853, China (Zhang ZB, Cao DS); Laboratory of Computational Physiology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States (Celi LA); Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02138, United States (Celi LA)

Corresponding author: Zhang Zhengbo, Email: zhengbozhang@126.com

【Abstract】 Medical practice generates and stores immense amounts of clinical process data, while integrating and utilization of these data requires interdisciplinary cooperation together with novel models and methods to further promote applications of medical big data and research of artificial intelligence. A "Datathon" model is a novel event of data analysis and is typically organized as intense, short-duration, competitions in which participants with various knowledge and skills cooperate to address clinical questions based on "real world" data. This article introduces the origin of Datathon, organization of the events and relevant practice. The Datathon approach provides innovative solutions to promote cross-disciplinary collaboration and new methods for conducting research of big data in healthcare. It also offers insight into teaming up multi-expertise experts to investigate relevant clinical questions and further accelerate the application of medical big data.

【Key words】 Big data in healthcare; Application of medical big data; Cross-disciplinary collaboration; Datathon

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)

医疗实践活动产生大量的数据,随着信息化的发展,我们得以用电子化的手段及时存储和记录这些医疗数据,对其进行深度挖掘利用对于提高医疗质量和患者安全具有重要

意义。随着数据量的累积以及大样本临床数据的建设,我们逐渐步入了医疗大数据以及人工智能时代。在医疗大数据时代,如何高效、有意义地挖掘利用这些数据,为临床决策

提供循证依据并发展智能化辅助决策支持工具呢?临床医生往往每天都面对着繁忙的诊疗任务,而数据科学家和工程师缺乏对临床知识的深入了解,因此,由任何一方独立开展医疗大数据应用研究都困难重重。医疗大数据的分析利用迫切地需要跨专业合作,组建一个理-工-医多学科团队,发挥各自专业优势,分工协作,实现共赢多赢。Datathon是医疗大数据时代的一种新兴跨学科合作模式,即基于“真实世界”的医疗数据,以患者和疾病诊疗为中心,通过组建跨学科团队合作的方式共同解答医疗实践中遇到的问题。Datathon模式在国外已实践多次,成功促进了多学科间的合作,推动了医疗大数据应用实践的落地。在国内,囿于长期以来学科之间缺乏交流合作以及缺乏信息、数据和成果共享的机制,跨学科的合作模式一直没有广泛建立起来。现通过介绍Datathon模式产生的时代背景、运作模式以及我们团队在Datathon模式方面的探索,将这种先进的数据实践模式引入国内,有助于我们尽快建立跨学科的团队协作,搭上医疗大数据和人工智能的快车,推动医疗大数据实践应用的落地。

1 跨学科合作模式的探索

1.1 竞争还是合作:在信息化时代,数据是丰富的研究资源,尤其是在医疗领域,医疗实践产生了海量的数据,这些数据包含丰富的患者信息,对这些数据进行深度挖掘利用对于提高医疗质量和患者安全具有重要意义。电子健康档案(EHR)是结构化的电子病历,包含了丰富的临床信息,即患者的疾病史、人口统计学特征、生理指标、生化指标、治疗干预、医嘱和护理记录等,记录了患者病情的演变过程以及这个过程中患者所接受的各种治疗,为我们理解疾病的发生发展过程,以及实现疾病的早期诊断、预测预警、临床决策支持和个性化医疗提供了丰富的数据集。目前已经有许多大型公开的EHR数据集,例如:麻省理工学院、飞利浦医疗系统以及贝斯以色列女执事医学中心共同合作发布的重症监护医学信息数据库(MIMIC)^[1],包含了超过4万患者的临床数据;美国患者结局研究机构(PCORI)主导整合的美国患者结局研究网络(PCORnet)数据库^[2],纳入了9个机构的医疗数据及个人分享的医疗数据;英国国家卫生服务体系(NHS)建立的开放的英国国家卫生服务(Open NHS)数据库^[3],包含了英国全部人口终生的医疗记录;飞利浦与麻省理工学院合作发布的eICU数据库^[4],包含了超过20万例患者的详细临床数据。这些数据集的开放使用为全世界的科研人员、临床医生、工程师、数据科学家提供了大样本的来自“真实世界”的研究资源集,客观上推动了相关领域的医学技术进步^[5]。

如何建设高质量的EHR数据库并开展EHR二次分析利用呢?虽然EHR有巨大的潜力,很多研究者却望而却步,原因在于操作EHR数据有其特殊性:首先,理解医疗数据需要丰富的专业知识,制定研究问题需要研究者对临床场景和传统的临床研究有一定了解,否则就会面对海量的医疗数据而无从下手,脱离医疗背景进行EHR研究是不可能

的;其次,EHR结构高度复杂,EHR数据的提取和整理需要研究者有一定的数据操作基础;另外,在数据分析层面,很多传统的数据分析方法不适用于大样本数据集,随着机器学习与人工智能的发展,数据分析的手段越来越多,将医疗大数据与新的人工智能方法结合可以帮助我们解决以往研究中的难题,而这些方法的运用需要研究者有一定的理工科背景。但现实是,临床医生往往缺乏数据库操作与大数据分析基础,并且每天面对繁忙的临床工作,而数据科学家和工程师缺乏对临床知识的深入了解,因此,由任何一方独立开发利用EHR数据都十分困难。当前学术界的竞争十分激烈,研究者们为了争取经费和发表文章,都希望独占数据资源,成为发现某项成果的“第一人”。但面对EHR这样大型的数据资源,以单打独斗的竞争模式进行研究非常困难。因此,挖掘和利用EHR数据迫切需要临床医生、工程师与其他学科的研究人员通力合作,进行跨学科的团队合作,才能有效应对医疗大数据时代的挑战,把握时代机遇^[6]。

1.2 Hackathon模式的兴起:Hackathon是近年来国外新兴的一种团队合作竞争模式,通常是以短期高强度的项目竞赛形式共同完成具有挑战的任务。这一模式最初起源于硅谷^[7],旨在集合工程师们在短期内共同开发软件。20世纪以来,这一模式在计算机领域被广泛应用,有不少公司诞生于Hackathon的项目中,如GroupMe、PhoneGap等等。这种竞赛机制能够促进团队合作,以头脑风暴的模式激发团队的执行力和创造力。在这个过程中,并不是仅有出色的想法就可以获胜,而是要有能在非常短的时间内将想法实现的执行力,团队内每个人的职责互补,大大提高了执行效率。

1.3 从Hackathon到Datathon:麻省理工学院和哈佛医学院团队将Hackathon模式引入医疗领域,提出了Data+Hackathon的竞赛模式,即“Datathon”^[8]。Datathon旨在汇集不同领域的从业人员,如临床医生、工程师、数据科学家等,通过组建跨学科的合作团队,基于数据和计算,在短期内协作完成临床研究项目,解答医疗领域中遇到的问题。在竞赛中,每个小组由不同领域的参赛者组成,分工合作,例如:临床问题由临床医生把控,技术层面的问题由工程师处理,数据分析中的问题由数据科学家和统计学家解决。在为期2d的竞赛后,每个小组对自己的项目进行汇报展示,并由各领域专家进行评审。这种跨学科的合作模式能够达成参与者之间的知识互补,打破既往学科间存在的壁垒,激发团队创造力。

自2014年起,麻省理工学院团队已经成功举办了19届Datathon活动。越来越多的人通过Datathon实践认识了医疗大数据,通过医疗大数据手段解决了临床问题。这一模式为医疗领域的跨学科合作提供了模版,并有很多竞赛小组实现了后续的团队合作。越来越多的EHR数据库向研究者开放,数据库中包含的数据越来越丰富详尽,这些数据库为研究者提供了宝贵的资源。面对丰富的大数据资源,各领域之间只有通过合作共享才能更高效地产出成果。Datathon的竞争合作模式为越来越多的参与者提供了跨学科交流合

作的机会,为推动医疗大数据应用落地提供了新思路。

1.4 国内跨学科合作模式的探索:国内 EHR 的发展起步较晚,质量参差不齐,对数据进行结构化十分困难,再加上数据共享的机制不成熟,各个机构并不愿意公开自己的数据资源。一方面,当前研究领域面对的竞争日益激烈,研究者们不愿意共享自己的研究数据;另一方面,数据的安全性越来越受到人们的重视,跨机构的数据共享存在诸多数据安全的顾虑。因此长期以来,各个研究机构间形成了独占数据资源进行研究的传统,在同一领域间尚且难以实现合作共享,跨学科合作更是难上加难^[5]。在倡导合作共赢的文化冲击下,“闭门造车”的科研文化早已不适应时代的发展。

2 Datathon 数据实践

近年来,医疗大数据的应用受到多方关注。自 2015 年起,政府出台了一系列政策文件^[9],促进规范健康医疗大数据融合共享与开放应用,推进健康中国战略。许多医院与企业开始积极探索如何应用医疗大数据^[9],发掘其中的价值。然而如何才能切实有效地开展医疗大数据应用研究,并推动项目落地,这亟需突破传统封闭的科研思路的桎梏,寻找新的思路方法,谋求合作共赢。

2016 年 10 月,解放军总医院联合麻省理工学院团队在北京举办了第一届医疗大数据研讨会与 Datathon 活动^[10]。2017 年 11 月成功举办了第二届 Datathon 活动^[11],有超过 1000 人参与了首日的研讨会,来自不同领域近 200 位参赛者组成了 15 个跨学科团队进行了 Datathon 数据实践。每个团队从选题论证开始,共同制定方案,分工提取数据,分析处理,最后得出结论,并于大会最后 1 d 分组进行了展示汇报。参与竞赛的临床问题主要涵盖重症监护科学、心脏内科学、急诊科学等领域,参赛人员根据提供的两个大型公开的数据库(MIMIC 数据库及 eICU 数据库)进行研究。活动的在线直播平台累计有超过 12 万人次的点击量,在业内引起了广泛关注和积极响应。2018 年 3 月,解放军总医院主持发布了解放军总医院急救数据库,并举办了第一届急救大数据 Datathon 活动,这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 人参与。这届“本土”Datathon 在临床数据库建设和共享利用方面进行了大胆的探索与尝试,进一步推动了我国临床数据库的开放共享和落地应用。

3 届 Datathon 活动的举办都非常成功,多个跨学科团队在短期内合作完成了感兴趣的临床课题,并有多个团队在活动后依然保持紧密合作,以论文的形式实现了项目产出。这些活动的实践成功为其他研究团队提供了参考范本,在用 EHR 进行的临床问题研究中,跨学科的合作模式能够发挥不同学科的优势,促进不同领域研究者的交流学习,在团队合作中实现多方审查,产生更可靠的研究结果。

3 经验总结与展望

医疗数据是大数据时代医学创新发展的基石,理-工-医跨学科进行科研合作是未来精准医疗、智慧医疗发展的

必然趋势。在信息爆炸、技术爆炸的今天,只有合作、良性竞争才能实现共赢。Datathon 模式为国内医疗领域的跨学科合作提供了创新协作的交流平台,为高效利用 EHR 提供了新的思路,这种合作竞争的方式能高效地促进不同领域从业者的交流协作,并通过短期的协作进一步促进建立长期合作,实现更长久的协同发展。

这一实践的成功得到了多方的广泛关注。在医疗行业,这是我们首次将呼吁“跨学科合作”从理念变成了现实,并且激发了多方探索数据库建设与数据共享的热情。高质量数据库的建设是医疗大数据分析的前提,我国在医疗数据体量上具有得天独厚的优势,只有做好 EHR 结构化、规范化,才能将数据转化为可用的资源。在此基础上,借助大数据与人工智能的结合,不断提高医疗系统的效能,构建基于数据驱动的学习型医疗系统(LHS)^[12],实现数据库建设-分析-应用的闭环。

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] Fleurence RL, Curtis LH, Califf RM, et al. Launching PCORnet, a national patient-centered clinical research network [J]. *J Am Med Inform Assoc*, 2014, 21 (4): 578-582. DOI: 10.1136/amiajnl-2014-002747.
- [3] Califf RM. The Patient-Centered Outcomes Research Network: a national infrastructure for comparative effectiveness research [J]. *N C Med J*, 2014, 75 (3): 204-210. DOI: 10.18043/nem.75.3.204.
- [4] Johnson AEW, Pollard TJ, Celi LA, et al. Analyzing the eICU collaborative research database [C]// The ACM International Conference. New York: ACM, 2017: 631. DOI: 10.1145/3107411.3107494.
- [5] Poucke SV, Gayle AA, Vukicevic M. Secondary analysis of electronic health records [M/OL]. Switzerland: Springer International Publishing, 2016 [2018-04-04]. DOI: 10.1007/978-3-319-43742-2.
- [6] Moseley ET, Hsu DJ, Stone DJ, et al. Beyond open big data: addressing unreliable research [J]. *J Med Internet Res*, 2014, 16 (11): e259. DOI: 10.2196/jmir.3871.
- [7] Leckart S. The Hackathon is on: pitching and programming the next killer app [EB/OL]. (2012-02-17) [2018-04-09].
- [8] 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.
- [9] 孟群, 毕丹, 张一鸣, 等. 健康医疗大数据的发展现状与应用模式研究 [J]. *中国卫生信息管理杂志*, 2016, 13 (6): 547-552. DOI: 10.3969/j.issn.1672-5166.2016.06.01.
Meng Q, Bi D, Zhang YM, et al. Research on the development and application of big data in healthcare [J]. *Chin J Health Inf Mana*, 2016, 13 (6): 547-552. DOI: 10.3969/j.issn.1672-5166.2016.06.01.
- [10] 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.
- [11] 曹德森, 张政波, 梁洪. 解放军总医院-麻省理工学院医疗大数据研讨会举行 [EB/OL]. (2017-11-23) [2018-04-09].
Cao DS, Zhang ZB, Liang H. The 2nd Chinese PLA General Hospital-MIT Health Data Conference and Workshop [EB/OL]. (2017-11-13) [2018-04-09].
- [12] Budronis 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.

(收稿日期: 2018-04-23)