

大数据与精准医学

Daniel Richard Leff and Guang-Zhong Yang*

摘要: 本文主要关注大数据分析对改善健康、预防和检测早期疾病以及个性化干预措施的潜在影响,讨论了大数据分析在建立患者电子健康档案以获得更好的临床决策支持中的可能作用。笔者研究了药物遗传学的发展,以更好地理解患者对化学疗法产生不同反应的原因。此外,笔者评估了在线健康交流的发展,以及如何利用这些数据检测公共健康威胁和控制传染病。最后,笔者描述了新一代可穿戴和可植入的躯体传感器如何改善身体状况、简化慢性病管理和提高外科植入物质量。

关键词: 大数据,生物传感器,人体传感网络,可植入传感器,临床决策支持系统,药物遗传学,移动医疗

1 引言

医疗保健领域产生数据的复杂性、多样性和丰富性正在不断推进与健康相关的大数据的发展[1]。大量、快速、多样、精准、变化和价值是大数据的显著特征,这些特征存在于生物医学和医疗信息学的固有挑战中。有效地应对这些挑战能够为着重于预防疾病、早期诊断和个性化治疗的更智能的医疗保健系统铺路。

2 适用于精准医学的大数据

患者的电子健康档案(EHR)是包含有关社会人口统计学、医疗状况、遗传学和治疗等信息的大数据来源,然而在缺少有效决策支持的情况下,人们处理这些数据的能力是有限的。对于医疗保健来说,大数据的目标是提供一套具备实时知识产出能力的持续学习型基础设施,并开发具有预防性、预测性和可参与性的医疗系统[2]。为了达到这一目标,临床医生需要利用计算机模型来帮助组织数据、识别模式、解释结果和设置行动阈值。利

用大数据分析产生新知识、改善临床护理、改进公共卫生监控的例子已有很多。例如,电子健康档案已经成功应用于药物的后市场监管及提高药物警戒。

在英国,国民健康服务计划将在2018年实现无纸化。电子健康档案将为未来的临床决策支持系统(CDSS)提供完整的数据来源,以弥补人们在数据理解和多任务处理上的缺陷。临床决策支持系统目前已在开发过程中,它可以获取并改进相关协议[3]、提醒服药时间[4]、提高筛查质量[5]以及预测再次入院[6]。当在特定的条件下有多种治疗方案时,临床决策支持系统会帮助临床医生选择最优方案。例如,治疗与前交叉韧带(ACL)有关的疾病的方案有物理治疗、药物治疗和手术治疗等。在这种情况下,分层研究患者的电子健康档案以及职业数据、休养数据和骨骼与肌肉状况的数据已被用于治疗方案的分类[7]。

人们期待健康大数据能够在药物遗传学和分层医疗保健领域发挥重要作用。相似癌症亚型的患者通常对同一种化学疗法(以下简称“化疗”)产生不同反应。例如,*CYP2D6*是与他莫昔芬应答有关的多态性基因[8],*BRAF*突变(*Y472C*)与达沙替尼在非小细胞肺癌中的应答有关[9],多基因标记近来被发现与直肠癌对同步放化疗的应答有关[10]。基因组的不稳定性被认为是观察到的药物反应多样性的原因。最近的研究致力于揭示基因组和化疗敏感性以及抵抗性和毒性之间复杂的相互作用[11-13]。例如,癌症基因组图谱(Cancer Genome Atlas)研究网络发起了一项泛癌症项目以分析多种癌症型和不同癌症型中的分子畸变[11],从而使学者能够发现新的畸变。与之类似的,癌细胞系百科全书(Cancer Cell Line Encyclopedia)[12]和癌症药物敏感性基因组学(Genomics of Drug Sensitivity in Cancer)[13]等项目正在建设巨大的基因组数据库,从而在成百上千种癌细胞系中,

The Hamlyn Centre, South Kensington Campus, Imperial College London, London SW7 2AZ, UK

* Correspondence author. E-mail: g.z.yang@imperial.ac.uk

© The Author(s) 2015. Published by Engineering Sciences Press. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

英文原文: Engineering 2015, 1(3): 277-279

引用本文: Daniel Richard Leff, Guang-Zhong Yang. Big Data for Precision Medicine. *Engineering*, DOI 10.15302/J-ENG-2015075

特异性研究基因组的生物标记物和药物敏感性之间的关系。作为人类能够利用大型药物遗传学数据库预测药物敏感性的证据,最近的研究表明通过基因组数据和药物应答数据可以改善用于预测单个癌细胞系药物敏感性计算的算法[14]。进一步的工作还包括利用接受了化疗的患者的肿瘤应答与毒性来验证该算法的预测结果。

社交媒体带来的通信的激增是医疗保健产生巨大变化的主要推动力之一。最近一项评估表明有十亿多条关于医疗保健的推文,这反映出医疗保健的提供者、患者、组织机构和各种第三方之间存在大量对话。据估计,约有20%的慢性病患者,如糖尿病、心血管疾病和癌症患者,会主动上网寻找其他病友并与之交流经验;这种行为催生了社交媒体上(如推特和脸书)患者群体的产生。社交媒体为研究者招募患者参与研究,以及为患者参与他们的健康数据共享提供了新的路径。例如,TuDiabetes和波士顿儿童医院(Boston Children Hospital)发起一个联合倡议——TuAnalyze [15],旨在使糖尿病患者在积极参与糖尿病研究的同时进行糖尿病监测、评估并共享他们的结果。大数据分析的一个最有趣的应用是能够预测和追踪主要疫情,从而增加公共卫生资源并利用社交媒体促进医疗保健信息向患者传播。通过社交媒体分析和环境数据的联合模型,可以更好地预测哮喘恶化等严重的医疗紧急情况。与预测滞后两周的传统模型不同,Ram等利用推特活动、谷歌搜索和空气质量数据成功建立了每日急性哮喘患者急诊入院量(入院量定义为低等、中等和高等)的精确预测模型[16]。同样,对于国际重大的公共卫生问题如埃博拉来说,上述联合模型可以收集并分析数据,从而支持疾病流行趋势早期预警系统的运行,并传播医疗保健教育干预信息[17]。例如,Odlum等通过分析关于埃博拉病毒检测(EVD)的推特活动发现至少在新闻提醒的三天前和官方疾控中心警告的七天前,推特上关于尼日利亚埃博拉确诊病例讨论的数量就已经开始逐步增加[17]。最后,许多学者正在研究社交媒体改变行为和改善医疗保健的潜力,包括以促进全球健康为目的的针对发展中国家的干预措施[18]。

大数据的另一个重要推动力是广泛分布的传感技术。由于技术进步,人们对可穿戴和可植入传感器的兴趣愈发浓厚,包括提升了带宽并改进了微电子技术的增强的无线通讯传感器。这些技术进步的结果使持续性、多模式和具有情境感知能力的传感器在制作上具有可行性[19]。与此同时,传感器在小型化、具体化、减轻生物污染和微电子制造等方面的进步提升了可植入生物传感器的多功能性和可靠性。因此,在面向个性化医疗保健与分层医疗保健的整合医疗中,片段式监测已被连

续感应和平行改进所取代。急性事件的高危患者,如心肌梗塞后引发的心律失常,可从血压、脉搏和心率的持续监测中受益,这样一来,心律失常可基本实现实时发现,并且相关信号可被发送到智能手机中进行下一步处理[20]。类似地,慢性疾病的患者,如糖尿病,可受益于能够持续监测的可植入微系统[21],从而进行长期血糖控制。可植入脊髓刺激器等神经调控装置可对慢性背部疼痛的控制产生影响[22]。安装在研究对象脊柱轴上的用于监测轴向负载的可植入传感器可能导致个性化骨科假体的出现[23]。未来,人们可以期待安装在手术干预部位附近用来监测局部血液动力学和组织离子含量的传感器能够发现伤口感染并且在亚临床阶段进行治疗,从而预防突发性败血症。我们还可以持续监测正在接受化疗的患者的白细胞和中性粒细胞计数,一旦出现中性粒细胞减少症,可以尽早用粒细胞刺激法进行处理,从而减轻败血症症状。

3 结论

总的来说,大数据的来源和计算技术正在快速发展,利用其改善健康水平和身体状况的可能方法也多种多样。对于医疗保健来说,大数据的目标是提供一套具备实时知识产出能力的持续学习型基础设施,并开发具有预防性、预测性和可参与性的医疗系统。显而易见,大数据分析在提高医疗保健和改善人口健康方面有巨大潜力。然而,若想成功地发挥这种潜力,还需要解决与数据隐私、安全、所有权和管理等有关的问题。

Compliance with ethics guidelines

Daniel Richard Leff and Guang-Zhong Yang declare that they have no conflict of interest or financial conflicts to disclose.

References

1. J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, G. Z. Yang. Big data for health. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1193–1208
2. L. Hood, M. Flores. A personal view on systems medicine and the emergence of proactive P4 medicine: Predictive, preventive, personalized and participatory. *New Biotechnol.*, 2012, 29(6): 613–624
3. J. G. Klann, V. Anand, S. M. Downs. Patient-tailored prioritization for a pediatric care decision support system through machine learning. *J. Am. Med. Inform. Assoc.*, 2013, 20(e2): e267–e274
4. B. G. Nair, S. F. Newman, G. N. Peterson, W. Y. Wu, H. A. Schwid. Feedback mechanisms including real-time electronic alerts to achieve near 100% timely prophylactic antibiotic administration in surgical cases. *Anesth. Analg.*, 2010, 111(5): 1293–1300

5. K. B. Wagholikar, et al. Clinical decision support with automated text processing for cervical cancer screening. *J. Am. Med. Inform. Assoc.*, 2012, 19(5): 833–839
6. J. Futoma, J. Morris, J. Lucas. A comparison of models for predicting early hospital readmissions. *J. Biomed. Inform.*, 2015, 56: 229–238
7. K. Mei, J. Peng, L. Gao, N. N. Zheng, J. Fan. Hierarchical classification of large-scale patient records for automatic treatment stratification. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1234–1245
8. M. P. Goetz, et al. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Tr.*, 2007, 101(1): 113–121
9. B. Sen, et al. Kinase-impaired BRAF mutations in lung cancer confer sensitivity to dasatinib. *Sci. Transl. Med.*, 2012, 4(136): 136ra70
10. T. Watanabe, T. Kobunai, T. Akiyoshi, K. Matsuda, S. Ishihara, K. Nozawa. Prediction of response to preoperative chemoradiotherapy in rectal cancer by using reverse transcriptase polymerase chain reaction analysis of four genes. *Dis. Colon Rectum*, 2014, 57(1): 23–31
11. Cancer Genome Atlas Research Network; J. N. Weinstein, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, 2013, 45(10): 1113–1120
12. J. Barretina, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 2012, 483(7391): 603–607
13. M. J. Garnett, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 2012, 483(7391): 570–575
14. J. Sheng, F. Li, S. T. Wong. Optimal drug prediction from personal genomics profiles. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1264–1270
15. Anon. TuAnalyze is here! 2010-05-19. <http://www.tuidiabetes.org/forum/topics/tuanalyze-is-here>
16. S. Ram, W. Zhang, M. Williams, Y. Pengetnze. Predicting asthma-related emergency department visits using big data. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1216–1223
17. M. Odlum, S. Yoon. What can we learn about the Ebola outbreak from tweets? *Am. J. Infect. Control*, 2015, 43(6): 563–571
18. S. Bahkali, N. Alkharjy, M. Alowairy, M. Househ, O. Da'ar, K. Alsurimi. A social media campaign to promote breastfeeding among Saudi women: A web-based survey study. *Stud. Health Technol. Inform.*, 2015, 213: 247–250
19. G. Z. Yang. *Body Sensor Networks*. 2nd ed. London: Springer-Verlag, 2014
20. F. Rincón, P. R. Grassi, N. Khaled, D. Atienza, D. Sciuto. Automated real-time atrial fibrillation detection on a wearable wireless sensor platform. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, NJ: IEEE Service Center, 2012: 2472–2745
21. M. M. Ahmadi, G. A. Jullien. A wireless-implantable microsystem for continuous blood glucose monitoring. *IEEE Trans. Biomed. Circuits Syst.*, 2009, 3(3): 169–180
22. J. Padwal, M. M. Georgy, B. A. Georgy. Spinal cord stimulators in an outpatient interventional neuroradiology practice. *J. Neurointerv. Surg.*, 2014, 6(9): 708–711
23. M. K. Moore, S. Fulop, M. Tabib-Azar, D. J. Hart. Piezoresistive pressure sensors in the measurement of intervertebral disc hydrostatic pressure. *Spine J.*, 2009, 9(12): 1030–1034