

医疗大数据应用 SWOT 分析

范月丹 潘 烁 李 俊 刘建炜

(中南大学信息安全与大数据研究院 长沙 410083)

[摘要] 介绍大数据的概念及在医疗领域的应用并进行 SWOT 分析, 其优势在于云计算可实现海量医学数据的存储、数据处理能力强大、基于数据分析的疾病预测具有前瞻性和时效性; 劣势在于医学数据收集难度大、数据共享差、整合困难、缺乏复合型人才; 挑战是患者隐私、数据缺乏及虚假数据等问题, 机遇是有国家政策支持、信息及传感技术发展、医学领域的巨大需求等。

[关键词] 大数据; 医疗卫生; 云计算; SWOT

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2016.07.012

SWOT Analysis of Medical Big Data Application FAN Yue-dan, PAN Shuo, LI Jun, LIU Jian-wei, Institute of Information Security and Big Data, Center South University, Changsha 410013, China

[Abstract] The paper introduces the concept of big data and its application in the medical field and makes SWOT analysis. The strengths include that the cloud computing can realize the storage of mass medical data, the data processing ability is powerful and the disease prediction based on data analysis is forward-looking and time-efficient. The weaknesses include the great difficulty in medical data collection, poor performance in data sharing, difficulty in integration, and lack of interdisciplinary talents. Regarding challenges and problems such as patients' privacy, data deficiency and false data, the opportunities include national policy support, development of information and sensor technology and huge demand in the medical field.

[Keywords] Big data; Healthcare; Cloud computing; SWOT

1 引言

近年来大数据越来越受关注, 很多行业都意识到它蕴含着巨大的价值。奥巴马政府曾发布《大数据研究和发展倡议》(Big Data Research and Development Initiative), 意在提升利用数据集合获取知识的能力并斥资两亿美元以上^[1]。伴随数据挖掘、云计算等新兴技术的不断发展, 从大量繁杂冗余的数

据中提取有价值的知识也成为可能。大数据时代已经来临, 它的出现正在改变人们的工作和生活方式, 渗透到各行各业中, 改变其运行模式。KDnuggets^[2]比较分析了 2015 年和 2014 年发布的新兴技术成熟曲线报告, 发现在 2015 年新兴技术成熟曲线中已经没有大数据的踪影, 意味着大数据失去了新兴技术的地位, 更趋向于实用化。大数据相关的关键技术, 如云计算、自动驾驶汽车、物联网、机器学习、市民数据科学家等都得到了发展。医疗卫生信息平台、医院信息化管理系统、数字化医疗器械在医疗卫生机构迅速普及, 产生了海量的医疗数据。将这些医疗数据进行合理化的利用, 更好地作

[修回日期] 2016-03-10

[作者简介] 范月丹, 在读本科生; 通讯作者: 刘建炜。

用于医疗事业很有价值，也是万众瞩目的大事。

2 大数据概论

2.1 定义

大数据时代预言家维克托，著名的数据科学家如此定义：大数据是当今社会所独有的一种新型能力，它以一种前所未有的方式，通过对海量数据进行分析，获得巨大价值的产品和服务^[3]。一般认为大数据是指需要新处理模式才能具有更强的决策力、洞察力和流程优化能力的海量、高增长率和多样化的信息资产^[4]。

2.2 意义

大数据的研究已经不止局限于科学研究及商业应用，很多国家已将其上升至国家战略层面，是评价国家综合实力的重要因素。2012 年美国启动“大数据发展计划”，每年投资 2 亿美元拉动大数据相关产业发展。该计划被视为美国政府信息科学领域的重大举措之一^[5]。我国政府也十分重视大数据的研究，“十二五国家科技计划信息技术领域备选项项目征集指南”更是将大数据研究列为首位。

3 大数据在医疗领域的应用

3.1 临床决策支持

电子病历的普及是医院信息化的重要表现，电子病历所涉及的医疗信息是丰富的医疗资源，利用大数据技术对其进行处理分析，可以从中提取重要的信息，发现有效的临床途径，尤其是临床决策支持系统可以为临床医生提供大量的医学支持，从而帮助临床医生做出最合理的诊断、选择最佳的治疗措施、提供最佳诊疗建议，同时也节约了医疗资源^[6-9]。Cichosz 等^[10]基于医学决策分析技术建立糖尿病及其并发症的模型，以提高对糖尿病患者的护理。Ko 团队^[11]利用大数据技术预测运动神经元疾病的严重程度，研究者基于 HBase 和 Apache Mahout 随机森林分类器分析 ALS 临床试验数据库提供的病患医疗记录信息，预测正确率达到 66%。另一

方面，临床决策支持系统的建立为医院高层决策提供重要依据。通过大数据的分析找到医院医疗质量不足和医疗资源分配不合理之处，从而帮助医院高层做出正确的决策，实现利益最大化^[9]。

3.2 个性化医疗

医疗卫生正在从以疾病为中心的医疗模式转变到以患者为中心的个性化医疗模式，同时基于个人基因的个性化诊疗越来越受到关注^[12]。经过几十年的研究，人们发现基因与多种疾病存在非常密切的联系，有时仅仅只是单基因的突变也能造成人类疾病或者特殊体质。通过对患者基因的测序判断疾病并制定治疗方案可提高效率。个性化医疗就是通过全面分析患者的包括基因组在内的生理数据集，联系疗效数据和特征数据，对特定疾病的易感性和对药物的特殊反应等，做出针对患者的特殊性治疗。乔布斯也曾对自身的 DNA 和肿瘤 DNA 进行排序，以制定治疗方案，得以延续生命，这也是个性化治疗的典型应用^[13]。Elbers 等^[14]提出个体动力学的概念，基于大数据技术，通过与电子病历直接互动，为每个病人在正确的时间提供抗生素的正确治疗剂量。英国为癌症患者建立了世界上最大的单一癌症数据库，该数据库可向研究人员及医生实时提供癌症数据，有助于推动癌症个性化治疗的进程^[15]。

3.3 药品研发

医药公司可以利用大数据提高研发效率。通过大数据分析技术对公众疾病的药物需求趋势进行分析，从而确定最有效率的投入产出比，以达到资源的最佳配置。通过对临床试验阶段或早期临床阶段的数据分析建立模型，及时预测药物的临床效果，降低研发成本。同时医药公司还可以缩短新药的研发周期，更快获得回报，医药公司可以将药物更快地推向市场，生产更有针对性、治愈率高、更有回报潜力的药物^[6]。谷歌研究发表题为“针对药物研发的大规模多任务网络”的文章，指出如何利用深度学习进行大数据的处理整合，以筛选出切实有效的药物成分，从而加快药物研发^[16]。医疗大数据分

析公司 Inovalon 称可凭借其强大的医疗数据集及基于云的分析技术，对药品研发的成本及经济效益等进行研究，以帮助制药公司制定最优的药品研发方案^[17]。

3.4 流行病预防与检测

大数据技术也应用于流行病的预防与检测。谷歌对流感的准确预测就是典型的例子。谷歌公司将高频检索词条与美国疾控中心的流感传播时期的数据进行比较，预测流感爆发，且判断十分及时^[18]。我国疾病预防控制中心建设的国家传染病与突发公共卫生事件网络直报系统已投入运行，且已经覆盖全国所有县级以上的疾控中心，每年有 600 多万个案信息上报。通过大数据分析技术，对全国的疫情起到很好的监测作用，以便及时采取有力措施控制流行病的扩散^[13]。有报道称研究人员利用计算机程序，以大数据技术拟合动物的生存模式，对某些人畜共患病做出预测，首先对黑热病、狂犬病等疾病的动物生存模式拟合，结果发现模型准确率达到 90% 以上，该程序已鉴定了可能携带人畜共患病的 150 种新物种，预测出 58 种新传染病^[19]。

4 大数据在医疗领域应用的 SWOT 模型分析

SWOT 分析方法是基于内外部竞争环境和竞争条件下的态势分析，S (strengths) 是优势、W (weaknesses) 是劣势，O (opportunities) 是机会、T (threats) 是威胁。SWOT 分析模型常被用于组织战略制定、竞争对手分析等^[20]。运用 SWOT 模型对医疗大数据应用进行分析，见图 1。

S	<ul style="list-style-type: none"> ❖ 云计算可实现海量数据的存储 ❖ 数据处理能力强 ❖ 分析预测能力 	W	<ul style="list-style-type: none"> ❖ 数据收集难度大 ❖ 数据共享性差 ❖ 数据整合困难 ❖ 人才瓶颈
T	<ul style="list-style-type: none"> ❖ 患者的隐私风险 ❖ 医疗数据缺失及虚假数据 	O	<ul style="list-style-type: none"> ❖ 国家政策大力支持 ❖ 解决医学难题的巨大需求 ❖ 信息及传感技术等的发展

图 1 大数据在医疗领域应用的 SWOT 矩阵

4.1 优势

4.1.1 云计算实现海量数据存储 大数据和云计算就像硬币的两面，密不可分。云计算具有面向大数据的存储、处理以及分析等技术和工具，为大数据的实际应用提供强大的技术支撑^[21]。美国国家标准与技术研究院 (National Institute of Standards and Technologi, NIST) 指出，云计算作为一种新型的计算模式，可通过互联网进行泛在接入，对资源进行按需访问及共享，使得服务更便捷，更具有弹性^[22-23]。云计算具体包含虚拟化、海量存储、并行计算等多项功能，可存储海量的数据。

4.1.2 数据处理能力强大 基于云的大数据处理技术和工具的发展，使得数据分析处理能力有了质的飞跃^[24]。谷歌开发的 BigQuery 产品、Microsoft 研发的 Hadoop on Azure 及 Amazon 的 EMR 技术都有大数据分析功能^[24-26]，为用户提供大数据处理的技术支持。

4.1.3 分析预测能力 通过对海量数据进行信息挖掘，可发现有价值的信息，如建立预测疾病模型，对流行病的控制及临床诊断提供及时、有价值的参考。也因为大数据强大的分析和预测功能，已经被应用于多个领域，如临床决策支持系统、个性化医疗、流行病预防与检测。在肿瘤预测研究中，Davoli 等^[27]将 3 大基因数据库基于云计算技术进行有效整合，建立 Tumor Suppressor and Oncogene 程序，将 8 200 种肿瘤配对基因信息进行综合分析处理，提出非整倍体不是癌症而是癌症驱动因素的观点。Steinberg^[28]研究团队利用大数据分析平台 Reverse Engineering and Forward Simulation (REFS) 对 36 944 例人群数据进行分析，前瞻性预测代谢综合征的发生风险，分析能力强，预测结果具有时效性。

4.2 劣势

4.2.1 数据收集难度大 医疗数据没有得到有效全面的收集。早期的医疗数据大多以纸质形式存储，缺乏电子形式，无法实现有效整合。现阶段，在很多县级以下的小医院、小诊所，电子病历也没有广泛普及。

4.2.2 数据共享性差 一方面表现为各医院之间信息得不到互通, 常见的情况是同一患者的病历数据及检查报告在不同的医院无法使用。另一方面, 医疗数据涵盖了电子病历、医疗保险、医学学术研究等多个范畴, 这些数据资源相对分散, 并不集中, 不利于实现共享^[29]。

4.2.3 数据整合困难 不同的医院采用的医院信息系统均不相同, 一定程度上使得系统的兼容性较差, 很难实现数据有效整合。

4.2.4 人才瓶颈 医疗大数据的专业人才不仅要掌握必要的医学知识, 更要具有计算机、网络、数据分析挖掘等知识。培养高质量的人才相对困难, 且目前开设相关专业的高校数目不多。

4.3 挑战

4.3.1 患者隐私风险 医疗大数据涉及个人健康信息的方方面面, 且大数据可对这些信息进行高度共享以及知识挖掘。个人健康信息作为个人隐私的重要部分, 是敏感和特殊的。大数据时代个人隐私保护面临着很大考验。美国 Target 公司曾对顾客的购买信息进行分析, 被投诉侵犯了他人的隐私权^[30]。美国医疗保险公司 Anthem 被盗走超过 8 000 万人的医疗信息^[31]。除了对个人隐私的简单泄露, 大数据技术使得对个人健康信息的预知成为可能, 其危险性不言而喻^[31]。

4.3.2 数据缺失及数据虚假 医疗大数据时代数据的缺失和虚假也是一大威胁。海量的医学数据在收集存储过程中难免会出现缺失或者错误, 一般是个人健康信息收集不全或者记录错误, 以及计算机系统误操作导致数据的不一致。数据不完整为之后的挖掘分析增加了不确定性, 而虚假数据可能导致预测结果的错误。中国食品药品管理总局对部分药物的临床试验机构进行核查时, 发现大部分存在临床试验数据不真实、不规范、不完整的现象, 8 家企业 11 种药品被指控临床数据造假, 而这只是冰山一角, 对药品的安全性及有效性判断流于形式^[32-33]。

大数据时代到来之际, 人们可能会过分夸大其作用, 盲目依赖大数据, 而忽视其局限性。尤其是

大数据在中国的发展还处于起步阶段, 很多数据的质量是否适于进行分析还有待考量, 曾有人对这些数据进行挖掘分析, 结果不客观, 导致预测结果有偏差, 进而误导决策, 带来很大的损失。

4.4 机遇

4.4.1 国家政策支持 近年来, 国家政策大力支持大数据的发展。《健康中国 2020 战略研究报告》中提出, 在未来要推进 7 大医疗体系建设^[34], 2014 年将医疗卫生信息化纳入“十三五”国家网络安全和信息化建设的重点^[1]。2015 年 9 月国务院发布了《促进大数据发展行动纲要》, 旨在于 2018 年底建成国家政府数据统一开放平台, 涵盖信用、交通、医疗等重要领域。此《纲要》从政策层面为医疗大数据的发展提供方向指引^[35]。

4.4.2 解决医学难题的供巨大需求 医学难题需要大数据技术提供解决方案。医学影像、传感以及智能机器人等技术的运用大大提高医疗诊断的质量和效率, 但还有很多医学难题亟待解决, 如肿瘤、心血管疾病、精神及神经性疾病以及一些复杂慢性病等正严重影响人们的生活质量及生命安全。这无疑给大数据的运用提供了很好的契机, 巨大的需求可促进大数据在医疗领域的实践。

4.4.3 信息及传感技术发展 信息技术和传感技术的不断发展, 移动终端不断普及为医疗大数据的发展提供支撑。埃里克·托普曾在他的著作中指出“随着传感技术、纳米技术等的发展, 对‘人’的信息感知, 已经打破了空间和时间的限制。医学诊断正在演化为全人全程的信息跟踪、预测预防和个性化治疗”^[36]。借助智能手机、云计算、无线传感等新兴技术, 人们可以实时监测自己的健康信息, 收集高质量的健康数据。如 Ario Health 公司开发的腕带有监测运动、心率睡眠等功能, 实时同步到移动终端^[37]。由麻省理工学院毕业生研制的“Mimo”连身衣可帮助家长实时监测婴儿的呼吸、体温、睡眠等数据, 在婴儿发生突发情况时可向家长发出紧急通知^[38]。

5 发展大数据技术需考虑的关键问题

5.1 健全法律法规

在大数据的冲击下,个人信息安全受到严重威胁。对个人隐私的泄露以及不正当利用事件屡有发生。建立并健全大数据相关的法律法规是大势所趋。中国已初步建立了个人隐私权保护的法律,但并不完整和全面,尤其在大数据的背景下,现有法律体系更显局限,在数据的收集、存储、处理等方面都未完善规则。美国发展大数据相对较早,已颁布了一系列关于大数据的法律法规,如 2012 年《大数据研究和发展计划》,以指导大数据的发展。我国可借鉴美国的经验,结合国情,构建大数据法律保障体系。

5.2 数据安全技术

数据安全技术是合理利用大数据的基础。采用先进的安全系统、加密技术、存储技术、访问控制等可以有效保护个人隐私。建议加大对大数据安全保障技术研发的资金投入,以引导科研人员不断自主创新,逐步发展大数据安全技术。

5.3 人才培养

医学大数据发展需要掌握多种学科知识和能力的复合型人才,且在医疗大数据的应用领域,对人才知识储备及能力要求更高。我国目前只有少数高校设立医学信息相关专业学科,也是造成人才匮乏的主要原因之一。建议在高校大力完善该学科建设,培养专业师资力量,加强医学信息专业学生在大数据方面的实践能力。此外将医学信息学相关课程作为医学生的通识教育,以培养医学生的医学信息素养。

6 结语

大数据时代给医疗卫生领域带来巨大的影响和推动。医疗健康与人类生存发展息息相关,大数据深入应用于医疗领域,将会产生巨大的经济及社会效益。在大数据快速发展的同时,也要意识到它的劣势及潜在威胁。希望在不久的将来,大数据能趋

利避害,在医疗领域最大程度地发挥作用。

参考文献

- 1 王忠. 美国推动大数据技术发展的战略价值及启示 [J]. 中国发展观察, 2012, (6): 44-45.
- 2 K Dnuggets. Gartner 2015 Hype Cycle: Big Data is Out, Machine Learning is in [EB/OL]. [2016-01-10]. <http://www.kdnuggets.com/2015/08/gartner-2015-hype-cycle-big-data-is-out-machine-learning-is-in.html>
- 3 维克托·迈尔·舍恩伯格, 肯尼斯·库克耶. 大数据时代: 生活、工作与思维的大变革 [M]. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013: 17.
- 4 柏秋云. 大数据的价值与挑战 [J]. 科技信息, 2013, (17): 479.
- 5 American Government. Big Data Across the Federal Government [EB/OL]. [2016-01-10]. http://www.whithous.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf.
- 6 张振, 周毅, 杜守洪, 等. 医疗大数据及其面临的机遇与挑战 [J]. 医学信息学杂志, 2014, (6): 2-8.
- 7 高汉松, 肖凌, 徐德玮. 基于云计算的医疗大数据挖掘平台 [J]. 医学信息学杂志, 2013, 34 (5): 7-12.
- 8 张怡, 李柯. 临床决策支持系统在医院的应用 [J]. 医学信息学杂志, 2015, (6): 27-30, 60.
- 9 姜奇平, 冯海超, 古福. 数据业务主营化大趋势 [J]. 互联网周刊, 2013, (7): 26, 46.
- 10 Cichosz SL, Johansen MD, Hejlesen O. Toward Big Data Analytics: Review of Predictive Models in Management of Diabetes and Its Complications [J]. J Diabetes Sci Technol, 2015, 10 (1): 27-34.
- 11 Ko KD, El-Ghazawi T, Kim D, et al. Predicting the severity of motor neuron disease progression using electronic health record data with a cloud computing Big Data approach [J]. IEEE Symp Comput Intell Bioinforma Comput Biol Proc, 2014, (5): 1-6.
- 12 黄新霖, 包小源, 俞国培, 等. 医疗大数据驱动的个性化医疗服务引擎研究 [J]. 中国数字医学, 2014, (8): 5-7.
- 13 张昌明, 朱红. 大数据及其在医疗领域的应用 [J]. 中国医学教育技术, 2015, (3): 294-297.
- 14 Elbers PW, Girbes A, Malbrain ML et al. Right Dose, Right Now: using big data to optimize antibiotic dosing in

- the critically ill [J]. *Anaesthesiol Intensive Ther*, 2015, 47 (5): 457-463.
- 15 英国建立世界最大单一癌症数据库 [J]. *中国信息界 - e 医疗*, 2013, (7): 23.
- 16 Bharath Ramsundar, Steven Kearne, Patrick Riley, et al. Massively Multitask Networks for Drug Discovery [J]. *Drug Discovery & Structural Biology*, 2015, (2): 1-27.
- 17 vcbeat. 医疗大数据分析公司 Inovalon 的上市之路 [EB/OL]. [2016-01-10]. <http://www.vcbeat.net/15695.html>.
- 18 相海泉. 迎接大数据时代 [J]. *中国信息界: e 医疗*, 2013, (5): 38-39.
- 19 Proceedings of the National Academy of Sciences. Could a computer predict the next pandemic? [EB/OL]. [2016-01-10]. <http://www.sciencemag.org/news/2015/05/could-computer-predict-next-pandemic>.
- 20 刘东冬, 刘胜林, 张佳华, 等. 基于 SWOT 模型分析大数据在医疗中的应用 [J]. *中国数字医学*, 2014, (11): 13-15, 18.
- 21 林文敏. 云环境下大数据服务及其关键技术研究 [D]. 南京: 南京大学, 2015.
- 22 P Mell, T Grance. The NIST definition of cloud computing [R]. Maryland: National Institute of Standards and Technology, 2009.
- 23 P. Mell and T. Grance. The NIST definition of cloud computing reconunendations of the national institute of standards and technology [R]. Maryland: National Institute of Standards and Technology, 2011.
- 24 J Tigani, S Naidu. Google BigQuery Analyrics [R]. New Jersey: John Wiley & Sons, 2014.
- 25 M Tulloch. Introducing Windows Azure for IT Professional [R]. Washington: Microsoft Press, 2013.
- 26 AWS Markeplsce [EB/OL]. [2016-01-10]. <https://aws.amazon.com/marketplace/>.
- 27 Davoli T, Xu A W, Mengwasser K E, et al. Cumulative Haploinsufficiency and Teiplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome [J]. *Cell*, 2013, (155): 948-962.
- 28 Steinberg G B, Church B W, McCall C J, et al. Novel Predictive Models for Metabolic Syndrome Risk: a "big data" analytic approach [J]. *Am J Manag Care*, 2014, (20): e221-228.
- 29 孙慧. 大数据在医疗行业中的应用与挑战 [J]. *解放军医院管理杂志*, 2015, (11): 1095-1096.
- 30 黄尤江, 贺莲, 苏焕群, 等. 医疗大数据的应用及其隐私保护 [J]. *中华医学图书情报杂志*, 2015, (9): 43-45.
- 31 刘星, 王晓敏. 医疗大数据建设中的伦理问题 [J]. *伦理学研究*, 2015, (6): 119-122.
- 32 中央政府门户网站. 药物临床试验数据造假 问题出在哪儿 [EB/OL]. [2016-01-10]. http://www.gov.cn/xinwen/2016-01/05/content_5030648.htm.
- 33 南方周末. 新药试验数据造假秘密 [EB/OL]. [2016-01-10]. <http://www.infzm.com/content/113117>.
- 34 朱彦, 徐俊, 朱玲, 等. 主要发达国家医疗健康大数据政策分析 [J]. *中华医学图书情报杂志*, 2015, (10): 13-17, 59.
- 35 国务院印发大数据纲要 涵盖医疗领域 [J]. *中国医院院长*, 2015, (19): 20.
- 36 [美] 埃里克·托普. 颠覆医疗: 大数据时代的个人健康革命 [M]. 北京: 电子工业出版社, 2014.
- 37 树里. 大数据掀起的医疗健康革命 [J]. *健康管理*, 2014, (7): 38-39.
- 38 39 健康网. 用大数据去养育宝宝: 智能监测器 [EB/OL]. [2016-01-10]. <http://baby.39.net/yeyp/my-cp/44/index.html>.