# **大数据："互联网+"医疗的基础 五大应用场景**

|  |
| --- |
| IMG_256  2 |
|  |

****医疗大数据的基本概念****

　　搭上“互联网+”东风的医疗，要走上发展快车道，离不开“大数据”这把金钥匙。如果我们把“互联网+”医疗比作一幢高楼大厦，那么，“大数据”就是这幢大厦的地基，万丈高楼平地起，没有坚固的地基，空中楼阁难触及；甚至，“大数据”还可以说是这幢大厦的砖瓦，没有充足的砖瓦，海市蜃楼不长久。一个小小的比方，让大数据的重要性不言而喻。这让我们在探究“互联网+”医疗之前，不得不先去认识一下如此风靡的“大数据”。

****大数据的前世今生****

　　回顾历史，“大数据”这个时髦的词汇，在物理学、生物学、环境生态学等领域，以及军事、金融、通讯等行业存在已有时日。但是真正引起人们高度关注的，还是因为近年来互联网和信息行业的迅速发展所致。

　　且让我们来看看“大数据”在互联网行业的表现。

　　首先，它源于互联网公司在日常运营中生成、累积的用户网络行为数据。这些数据的规模是如此庞大，以至于人们将不能再用G或T来衡量，所以，大数据的起始计量单位至少是P（1000个T）、E（100万个T）或Z（10亿个T）。

　　那么，大数据到底有多大？一组名为“互联网上一天”的数据告诉我们：一天之中，互联网产生的全部内容可以刻满1.68亿张DVD；发出的邮件有2940亿封之多（相当于美国两年的纸质信件数量）；发出的社区帖子达200万个（相当于《时代》杂志770年的文字量）；卖出的手机为37.8万台，高于全球每天出生的婴儿数量37.1万……

　　截止到2012年，数据量已经从TB（1024GB=1TB）级别跃升到PB（1024TB=1PB）、EB（1024PB=1EB）乃至ZB(1024EB=1ZB)级别。国际数据公司（IDC）的研究结果表明，2008年全球产生的数据量为0.49ZB，2009年的数据量为0.8ZB，2010年增长为1.2ZB，2011年的数量更是高达1.82ZB，相当于全球每人产生200GB以上的数据。而到2012年为止，人类生产的所有印刷材料的数据量是200PB，全人类历史上说过的所有话的数据量大约是5EB。IBM的研究称，整个人类文明所获得的全部数据中，有90%是过去两年内产生的。而到了2020年，全世界所产生的数据规模将达到今天的44倍。

　　一系列的数据无不向我们传达着一个概念，那就是“大”，以几何倍数增长的大，而且越来越大。但是，如果就此顾名思义，定义了“大数据”，那还是有失偏颇。

　　最早提出“大数据”时代到来的，是全球知名咨询公司麦肯锡，麦肯锡称：“数据，已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。”

　　牛津大学互联网研究所维克托·迈尔·舍恩伯格教授指出，“大数据”所代表的是当今社会所独有的一种新型的能力——一种前所未有的方式，通过对海量数据进行分析，获得有巨大价值的产品、服务和见解。

　　“大数据”会给整个社会带来从生活到思维上革命性的变化：企业和政府的管理人员在进行决策的时候，会出现从“经验即决策”到“数据辅助决策”再到“数据即决策”的变化；人们所接受的服务，将以数字化和个性化的方式呈现，借助3D打印技术和生物基金工程，零售业和医疗业亦将实现数字化和个性化的服务；以小规模实验、定性或半定量分析为主要手段的科学分支，如社会学、心理学、管理学等，将以向大规模定量化数据分析转型；将会出现数据运营商和数据市场，以数据和数据产品为对象，通过加工和交易数据获取商业价值；人类将在哲学层面上重新思考诸如“物质和信息谁更基础”、“生命的本质是什么”、“生命存在的最终形态是什么”等本体论问题。

　　综上所诉，我们对“大数据”概念基本能有个全方面的认识。“大数据”不是数据量的简单刻画，也不是特定算法、技术或商业模式上的发展，而是从数据量、数据形态和数据分析处理方式，到理念和形态上重大变革的总和。所谓“大数据”，是基于多源异构、跨域关联的海量数据分析所产生的决策流程、商业模式、科学范式、生活方式和观念形态上的颠覆性变化的总和。

****接轨“大数据”，传统医疗遇挑战****

　　搭上“互联网+”的列车，医疗机构便开始行进在医疗信息数字化的路上。

　　医疗数据是医生对患者诊疗和治疗过程总产生的数据，包括患者基本数据、电子病历、诊疗数据、医学影像数据、医学管理、经济数据、医疗设备和仪器数据等，以患者为中心，成为医疗信息的主要来源。而不断数据化的信息，在使医院数据库信息容量不断膨胀的同时，也对疾病及病人的管理、控制和医疗研究起到了积极的作用，价值不菲。

　　但是，发展了那么多年的传统医疗行业，现在要跨入“互联网+”的频道，其固有的复杂性和特殊性致使转型难度不容小觑。就拿一个初具规模的医院来说，每天需要接待上万的患者前来就诊，患者的基本信息、影像信息与其他特殊诊疗信息汇集在一起，那将是一个非常庞大的数据。

　　据初步统计，上海市区域医疗信息平台（上海市“医联工程”及区县卫生数据中心）已经积累了覆盖3900万人群、1400 TB数据量的电子诊疗与健康档案等医疗卫生数据（涵盖了全市38家三级医院3900万就诊人群的诊疗信息，包括患者基本信息、就诊信息、健康档案、检验及影像检查报告、医学影像图像文件、住院相关病历、医保结算等医疗卫生数据，涉及就诊记录2.1亿条，处方记录9.1亿条）。

　　日积月累，这个数据量将会持续快速增长，为医院的数据存储、集成、调用等应用带来巨大压力。除了数据规模巨大之外，医疗行业的数据类型和结构极其复杂，如PACS影像、B超、病理分析等业务产生的非结构化数据，这些数据存储复杂，并且对传统的处理方法和技术带来巨大挑战。

****医疗大数据的主要来源****

　　随着医疗卫生信息化建设进程的不断加快，医疗数据的类型和规模也在以前所未有的速度迅猛增长，甚至到了在很大程度上无法利用目前主流软件工具，在合理的时间内达到撷取、管理并整合成为能够帮助医院进行更积极目的经营决策的有用信息的地步。而且，如此具有特殊性、复杂性的庞大的医疗大数据，其搜集如果仅靠个人甚至个别机构，那基本是不可能完成的任务。那么，这些数据到底是怎么产生的，又都来自于哪里呢？经过简单的梳理，我们大致可以把他们归档在以下4个框架里：

　　（1）病人就医过程中产生的信息。从患者进入医院开始，挂号环节便将个人姓名、年龄、住址、电话等信息输入完全了；随后在医生就医环节，病患的身体状况、医疗影像等信息也将被录入数据库；看病结束以后，患者买单结算的过程中，又将有费用信息、报销信息、医保使用情况等信息被添加到医院的大数据库里面。这将形成医疗大数据最基础却也是最庞大的原始资源。

　　（2）临床医疗研究和实验室数据。临床和实验室数据整合在一起，使得医疗机构面临的数据增长非常快，一张普通CT图像含有大约150 MB的数据，一个标准的病理图则接近5 GB。如果将这些数据量乘以人口数量和平均寿命，仅一个社区医院累积的数据量就可达数万亿字节甚至数千万亿字节(PB)之多。

　　（3）制药企业和生命科学。药物研发所产生的数据是相当密集的，对于中小型的企业也在百亿字节（TB）以上的。在生命科学领域，随着计算能力和基因测序能力逐步增加，美国哈佛医学院个人基因组项目负责人詹森·鲍比就认为，到2015年，将会有5000万人拥有个人基因图谱，而一个基因组序列文件大小约为750MB。

　　（4）智能穿戴设备带来的健康管理。随着移动设备和移动互联网的飞速发展，便携式的可穿戴医疗设备正在普及，个体健康信息都将可以直接连入互联网，由此将实现对个人健康数据随时随地的采集，而带来的数据信息量将更是不可估量的。

****医疗大数据的特性：大数据性+医疗性****

　　如此规模巨大的临床实验数据、疾病诊断数据以及居民行为健康数据等汇聚在一起所形成的医疗大数据，已然呈现出其作为大数据的特性，即：

　　(1)数据规模大(volume)。例如一个CT图像含有大约150MB的数据，而一个基因组序列文件大小约为750MB，一个标准的病理图则大得多，接近5GB。

　　(2)数据结构多样(variety)。医疗数据通常会包含各种结构化表、非（半）结构化文本文档(XML和叙述文本)、医疗影像等多种多样的数据存储形式。

　　(3)数据增长快速(velocity)。一方面，医疗信息服务中包含大量在线或实时数据分析处理，例如，临床决策支持中的诊断和用药建议、流行病分析报表生成、健康指标预警等；另一方面，得益于信息技术的发展，越来越多的医疗信息被数字化，因此在很长一段时间里，医疗卫生领域数据的增长速度将依然会很快。

　　(4)数据价值巨大(value)。毋庸置疑，数据是石油，是资源，是资产，医疗大数据不仅与每个人的个人生活息息相关，对这些数据的有效利用更关系到国家乃至全球的疾病防控、新药品研发和顽疾攻克的能力。

　　而除了大数据所具有的特征(即volume，variety，value，velocity)外，医疗大数据还具有多态性、不完整性、时间性及冗余性等医疗领域特有的一些特征。

　　(1)多态性医疗大数据包括纯数据（如体检、化验结果）、信号（如脑电信号、心电信号等）、图像（如B超、X线等）、文字（如主诉、现／往病史、过敏史、检测报告等），以及用以科普、咨询的动画、语音盒视频信息等多种形态的数据，是区别于其他领域数据的最显著特征。

　　(2)不完整性医疗数据的搜集和处理过程经常相互脱节，这使得医疗数据库不可能对任何疾病信息都能全面反映。大量数据来源于人工记录，导致数据记录的偏差和残缺，许多数据的表达、记录本身也具有不确定性，病例和病案尤为突出，这些都造成了医疗大数据的不完整性。

　　(3)时间性患者的就诊、疾病的发病过程在时间上有一个进度，医学检测的波形、图像都是时间函数，这些都具有一定的时序性。

　　(4)冗余性医学数据量大，每天都会产生大量信息，其中可能会包含重复、无关紧要甚至是相互矛盾的记录。

****医疗大数据的应用及案例****

　　经过数据的原始积累，并逐步走向成熟的医疗大数据，将给我们带来怎样的效用呢？根据全球管理咨询公司麦肯锡的一份最新报告显示，医疗保健领域如果能够充分有效地利用大数据资源，医疗机构和消费者便可节省高达4500亿美元的费用。

　　大数据在医疗行业的应用可在以下几个方面发挥积极作用：

　　（1）服务居民。居民健康指导服务系统，提供精准医疗、个性化健康保健指导，使居民能在医院、社区及线上的服务保持连续性。例如，提供心血管、癌症、高血压、糖尿病等慢性病干预、管理、健康预警及健康宣教（保健方案订阅、推送）；同时减少患者住院时间，减少急诊量，提高家庭护理比例和门诊医生预约量。

　　（2）服务医生。临床决策支持，如用药分析、药品不良反应、疾病并发症、治疗效果相关性分析、抗生素应用分析；或是制定个性化治疗方案。

　　（3）服务科研。包括疾病诊断与预测、提高临床试验设计的统计工具和算法、临床实验数据的分析与处理等方面，如针对重大疾病识别疾病易感基因、极端表现人群；提供最佳治疗途径。

　　（4）服务管理机构。规范性用药评价、管理绩效分析；流行病、急病等预防干预及措施评价；公众健康监测，付款（或定价）、临床路径的优化等。

　　（5）公众健康服务。包括危及健康因素的监控与预警、网络平台、社区服务等方面。

(责任编辑：UF023)原标题：大数据——“互联网+”医疗的基础，五大应用场景