

医疗大数据分析技术研究进展

黄 浩

浙江辰时科技集团有限公司 浙江 杭州 310000

摘 要：随着经济的迅速发展和信息技术的变革，大数据正在改变全球绝大部分行业，医疗领域也不例外。信息化时代下，在健康中国战略的推进中，应将大数据技术在未来医疗领域当中的应用，作为一项重大课题，积极加强研究、探讨与实践，一方面既要正确、深入认识大数据技术的概念及其在医疗领域中的作用优势，客观看待将大数据技术应用用于未来医疗领域的必要性，另一方面更要采取科学、合理的应用途径，将相关技术优势在医疗领域中切实地发挥出来，打造全新的、信息化的医疗卫生服务体系，提升医疗卫生服务质量水平。

关键词：医疗；大数据；技术

引言

大数据是一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合，具有海量的数据规模、快速的数据流转、多样的数据类型以及价值密度四大特征。本文针对医疗领域对大数据的运用进行分析。

1 医疗大数据的意义

医疗大数据的研究和应用有如下几方面的现实意义：（1）提高诊断准确性。医生利用医院共享的大数据系统，能够分析医院积累的大量临床数据，结合目标患者信息，能够及时、有效地为患者选择治疗方案；（2）减少医患矛盾。通过建立移动护理信息系统，实现病人护理文档的电子化，将医生、护士、其他医务人员工作闭环执行，使患者得到有效治疗；（3）节约医疗资源。通过云计算技术，将信息资源集成和共享，降低成本的同时提高效率；（4）个性化医疗。通过大数据挖掘技术，对个体差异性特征在普遍特征数据分析的基础上进行优化分析，进而辅助以医疗服务子单元，满足个性化医疗服务要求。

2 医疗大数据分析技术特征

2.1 规模性

大数据的特征首先就是数据规模大。随着互联网、物联网、移动互联技术的发展，人和事物的所有轨迹都可以被记录下来，数据呈现出爆发性增长。从现状来看，基本上是指从几十TB到几PB，这样的数量级。当然，随着技术的进步，这个数据量不断增大。这是由于互联网，物联网、传感器和各种智能设备的发展导致的。Google据统计每个月要处理900亿次的Web搜索，为此每月要处理的数据高达600PB。

2.2 多样性

数据来源的广泛性，决定了数据形式的多样性。大数据可以分为三类，（1）结构化数据，如财务系统数据、信息管理系统数据、医疗系统数据等，其特点是数据间因果关系强。（2）非结构化的数据，如视频、图片、音频等，其特点是数据间没有因果关系。（3）半结构化数据，如HTML文档、邮件、网页等，其特点是数据间的因果关系弱。Twitter和Facebook等社交媒体中的文本数据、智能手机中内置定位系统产生的定位信息、时刻生成的传感器数据，还有图片、文本、视频、音频等非结构化数据，这些充分体现出数据多样性的特点。近年来，非结构化数据增长迅速，这些数据都是用一般的关系数据库很难存储的，需要大数据的储存技术。

2.3 高速度

数据的增长速度和处理速度是大数据高速性的重要体现。在大数据时代，大数据的交换和传播主要是通过互联网和云计算等方式实现的，其生产和传播数据的速度是非常迅速的。另外，大数据还要求处理数据的响应速度要快。

2.4 价值性

尽管企业拥有大量数据，但是发挥价值的仅是其中非常小的部分。大数据背后潜藏的价值巨大。由于大数据中 valuable 的数据所占比例很小，而大数据真正的价值是从大量不相关的各种类型的数据中获取的。我们通过机器学习方法、人工智能方法或数据挖掘方法进行深度分析，可以从海量获取更大的价值。

3 大数据技术在医疗领域中的运用

3.1 医疗信息化建设

医疗信息化即医疗服务的数字化、网络化、信息化。早期的医疗记录的常见做法是以手写笔记或打印报告的形式，随着计算机系统的出现及其潜力，医疗保健

系统中所有临床检查和病历的数字化已成为当今的标准和广泛采用的做法。由Murphy、Hanken、Waters定义电子健康记录(ElectronicHealthRecords, EHR)是用于患者过去、现在或未来的身体/心理健康或状况有关的任何信息的计算机化医疗记录系统,方便医疗人员访问患者的医学诊断、处方、与已知过敏相关的数据,提高医疗识别和治疗效率。与之类似,电子病历(EMR)、个人健康记录(PHR)、医疗实践管理软件(MPM)和许多其他医疗数据系统共同为提高医疗质量、服务效率和降低成本以及减少医疗错误提供有力保障。

3.2 临床医护决策辅助

在传统的医疗诊断主要依靠医生个人经验和目标患者的信息,局限性很大,由于患者的个体差异,同样诊断治理方式应用在不同患者身上的效果也可能不同。为此,可借助大数据建立一个智能诊断系统,如果将所有患者的就诊信息整合,运用数据挖掘技术分析患者信息,提高诊断准确率。另外,可利用自然语言处理技术对电子病历文本进行分析,图像智能识别技术对医学影响建模,实现智能诊断。这能促进诊断更加科学化、合理化,提高诊断准确率,有效降低误诊率。

3.3 医疗科研领域中的技术支持

在医疗科研领域,包括治疗方案的疗效判断、新药研发、器械研制、疾病危险因素分析等各类临床科研。运用大数据技术对各种数据进行筛选、分析,可以为科研工作提供强有力的数据分析支持。如:(1)探索疾病关联关系。将离散数据进行整合与规范化,对大量、关联性的疾病数据进行分析整理,构建医疗知识图谱,可以探索疾病的关联关系;(2)进行临床预测。通过数据挖掘技术对各类疾病的症状,监控疾病治疗效果数据进行建模,预测处于早期症状的疾病的发展趋势。

3.4 公共卫生领域

运用大数据可以对流行病的爆发和规模进行预测预警。现阶段,人类处于流行病的高发期,运用大数据技术可以通过传播时期的监测数据(异常病例增加数,检索词条、微博论坛等),建立处理模型,从而达到预测预警的目的。大数据对流行病调查与追踪也有着重要作用。

3.5 医学影像的应用

现阶段,医院的存储容量也大部分被医学影像数据所占据,包括X射线,核磁共振成像,超声波成像等。这些都是非结构化的影像数据,而大数据的优势就是对非结构化数据的储存、分析、处理。大数据技术结合人工智能(例如神经网络)开发出来的能进行医学影像分析的“阅片机器人”,将成为放射科医师得力助手,将大

大缩短诊断时间。

4 大数据在医疗领域的挑战及应对

4.1 数据集成和共享

大型医院已经基本实施了信息化建设,但相应的数据资源仍然分散在不同的数据池中,互不相连,形成信息孤岛,导致医院之间无法共享护理数据,全社会卫生系统的相关数据无法共享和有效整合,进而影响医疗资源数据池的形成。遏止健康医疗数据共享的关键问题主要体现在管理机制和技术标准两个方面,因此亟需解决以下2个问题:(1)由于我国经济发展的地区性差异,不同地区和级别的医疗机构自成体系、自建标准,不同类型医疗机构间缺乏充分的跨业务数据共享,因此应该健全协作机制,调整利益冲突并通过管理手段达成共识;(2)医疗大数据本身多源异构性的特点,存在大量结构化、半结构化、非结构化数据,信息标准并不统一,按照标准规范建立起跨机构甚至跨区域的数据共享平台及数据交换机制。

4.2 数据安全和隐私保护

大数据在促进数据传输和共享的同时,也带来了新的挑战,数据隐私保护就是其中之一。我国对于医疗领域隐私保护的立法及政策法规的制定处于起步和发展中,尽管已有相关法律法规对保护患者隐私提出了相应的要求和规定,但是大部分规定缺乏可操作性,仅规定了对患者的保密义务,而没有规定违反该义务的后果,不利于具体司法实践操作,而且大多数条款对患者信息权利的具体内容、权利保护的方式等都没有规定。大数据时代,要搞数据共享,这在很大程度上受到严格的数据保护的制约;只有保障大数据的共享和保护,才能最大限度地发挥医疗大数据的潜在价值。为了能够最大程度上保护公民的隐私安全,需要加强对数据的保护工作。首先在法律的内容上,应尽快制定相关政策法规;除完善相关配套法规标准外,也要在执法和司法过程中强化保护患者隐私与个人信息,严格要求数据安全保障义务与责任,坚持知情同意原则,数据的收集、存储、使用、分析等行为应获得数据主体的知情同意,原则上应获得明示同意;其次在治理手段上,在强化隐私保护技术对个人隐私保护效果支撑作用的同时,致力提升数据加工者合法合规使用数据的自觉意识,形成注重数据伦理安全的良性氛围,倡导公民形成正确的个人隐私观。

4.3 数据质量保障

医疗数据在大部分应用场景中,对数据的质量都有比较高的要求。医疗数据质量风险可能导致一系列问题:(1)即时接受诊疗的患者因为医生收到了错误的

基本疾病信息受到不合理的治理甚至承担安全风险；

(2) 医疗科研有可能因为出现的错误数据而得到错误的结论,轻者会导致整个科研过程失败,耗费人力财力,重者可能会错误引导医学理论发展方向,影响整体医学发展进程。为了提高医疗数据的准确性、真实性、适用性,须建立一套标准化数据治理模式,各级医疗机构在开展医疗数据采集、存储、整合、分析等工作的时候,以相同或相近的标准,确保内容准确、术语规范、信息数据接口一致等。此外人工智能技术也正在应用于医疗数据的提取和识别过程中,通过自然语言处理、人工智能、OCR等技术直接识别电子病历系统并完成有效数据的提取,这一过程的实现将有效减少医疗数据提取过程的人力耗费且有可能通过机器学习提高数据的准确性。这些医疗数据提取的自动化策略必然是医疗数据提取的重要未来方案,对于解决当下的医疗数据准确性安全性风险意义重大。

4.4 医疗领域对大数据的运用策略

医疗领域对大数据的运用存在的问题主要在这几个方面:医院存储系统中数据存储分散,数据的完整性,一致性需要提高,而且需要进行一系列的数据清洗才能进行大数据分析;医疗方面的数据涉及个人及医院的隐私问题比较多,一旦泄露数据将造成患者歧视、信息非法流通等风险,也给电子病历的全面共享带来难度。这个需要在法律法规上寻求突破;医疗领域大数据应用人才队伍不足。医疗领域要对大数据充分运用,需要建设既有医学背景又掌握大数据的相关技术,这样的复合型

人才队伍。

结束语

当前,信息化技术已经得到了相当广泛的应用和普及,深入地渗透进去了各个行业、领域,而且除了互联网技术、计算机技术之外,信息化技术的内涵和种类也越来越多,如大数据技术,它们相互之间的联系紧密,将人类社会真正的带入到了信息化时代。在这样的背景下,应当意识到大数据技术在医疗领域当中,具有重要的作用价值,对健康中国的构建、满足群众生活习惯、促进医学研究、优化医疗卫生服务资源分配来说意义重大。为此,应当加强对大数据技术的探索和应用,结合实际的医疗卫生服务需求,将其功能、优势充分地发挥出来,打造面向未来的信息化医疗卫生服务体系,从而提升医疗卫生服务质量水平,予以人们更加优质、可靠的健康保障。

参考文献

- [1]何璋函.大数据技术在医疗领域的应用[J].电子元器件与信息技术,2022,6(03):54-55+58.
- [2]赵鑫.大数据与云计算在医疗系统中的应用[J].科学技术创新,2017(22):121-122.
- [3]王海星,张靓,杨志清,郭燕青,郭玲玲,齐昊.医疗大数据在临床科研中的应用探讨[J].中国医院,2020,24(07):63-64.
- [4]赵贇.基于大数据的医疗卫生信息平台建设探讨[J].数码世界,2018(08):143.