

大数据背景下信息系统在COVID-19疫情中的应用浅析

古雪颖¹, 尚志刚², 郭曦榕^{3*}

¹成都信息工程大学, 物流学院, 四川 成都

²电子科技大学, 信息与软件工程学院, 四川 成都

³成都信息工程大学, 管理学院, 四川 成都

Email: *szgtql@outlook.com

收稿日期: 2020年11月24日; 录用日期: 2020年12月9日; 发布日期: 2020年12月16日

摘要

随着大数据技术近年来在我国的蓬勃发展, 医学领域的大数据应用也日新月异。COVID-19疫情的爆发使得大数据技术在疫情管控过程中发挥了重要作用, 如疫情信息管理、疫情分析、疫情应急指挥等。信息系统作为大数据技术应用的基本载体, 在COVID-19疫情中得到了广泛的应用, 是大数据技术在疫情防控中与实际相结合的具体体现。本文在分析信息系统在COVID-19疫情管控中应用现状的基础上, 对当前应用中存在的问题进行了分析, 并提出了建议与展望。

关键词

信息系统, 大数据, 新冠病毒, 应用现状

Analysis of the Application of Information Systems in the COVID-19 Epidemic under the Background of Big Data

Xueying Gu¹, Zhigang Shang², Xirong Guo^{3*}

¹School of Logistics of Chengdu University of Information Technology, Chengdu Sichuan

²School of Information and Software Engineering of UESTC, Chengdu Sichuan

³School of Management of Chengdu University of Information Technology, Chengdu Sichuan

Email: *szgtql@outlook.com

Received: Nov. 24th, 2020; accepted: Dec. 9th, 2020; published: Dec. 16th, 2020

*通讯作者。

Abstract

With the vigorous development of big data technology in our country in recent years, the application of big data in the medical field is also changing day by day. The outbreak of the COVID-19 epidemic has enabled big data technology to play an important role in the process of epidemic management and control, such as epidemic information management, epidemic analysis, and emergency response command. As the basic carrier for the application of big data technology, information systems have been widely used in the COVID-19 epidemic and are a concrete manifestation of the combination of big data technology in epidemic prevention and control with reality. Based on the analysis of the status quo of the application of information systems in the management and control of the COVID-19 epidemic, this paper analyzes the existing problems in the current application, and puts forward suggestions and prospects.

Keywords

Information System, Big Data, COVID-19, Application Status

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当前大数据技术正在从深度和广度上应用到各行各业, 医疗和公共卫生领域也是大数据应用的重点领域, 如利用大数据进行疾病的诊断和预测、基因序列识别和健康管理等方面。新冠肺炎疫情爆发以来, 利用大数据实现了快速防疫和精准疫情控制。信息系统作为大数据应用的载体, 通过对数据进行采集、处理、分析、挖掘和可视化表达等过程, 面向不同的用户和需求为其提供专门的服务, 形成了多种功能多样、特点突出、直观实用的信息系统, 在疫情防控的过程中发挥了重要的作用。本文以 COVID-19 疫情为例, 分析了信息系统在疫情防控中的应用场景和现状, 并针对当前疫情防控信息系统中存在的问题提出了改进和发展建议。

2. 信息系统在 COVID-19 疫情中的应用现状

新冠疫情的防控用实践证明大数据在医疗信息系统中应用。通过对疫情期间数据的实时采集、处理、分析、挖掘再输入医疗信息系统, 加强疫情溯源和监测, 能更为有效地促进信息透明, 实时发布权威信息, 解决信息不对称的问题。信息系统在疫情信息管理、信息可视化、传播趋势预判、传播模型的构建、疫情早期趋势以及加快病毒检测诊断速度这几个方面都有显著的应用成果, 为疫情防控提供了更加全面的、自动化的管理以及各种服务, 加速实现了线上线下深度结合, 为信息公开、疫情监测、民生保障创造有利条件。

2.1. 疫情信息管理

疫情信息管理要解决的首要问题是“信息不通, 时间滞后”, 为此我国在 SARS 疫情后, 建立了世界上规模最大的传染病疫情突发公共卫生事件网络直报系统, 该系统不仅解决了信息自下而上的逐级传递问题, 并且大大提升了数据的汇总速度, 为抗击疫情争取到了宝贵的时间。信息管理能够提高了全社会的协作效率, 在大数据背景下, 信息管理系统将会成为跟战略应急物资一样的社会应急基础设施。

数据对于疫情防控具有重要的决策价值，因此数据的收集在信息管理中至关重要，涵盖了医院所提供的病例数据、以及人口流动、物资流动、交通、社区居民等的疫情相关数据。社区管理所采用的二维码线上采集数据是一个重要的方式，社区工作者可以快速收集追踪返乡人员信息，时间更快、效率更高、接触更少。

疫情信息的监控也是至关重要的，为应对 COVID-19 疫情所引发的公共卫生紧急状况，约翰·霍普金斯大学系统科学与工程中心推出了一个在线互动式仪表盘，用于跟踪全球的疫情。约翰·霍普金斯大学“仪表盘”更重要的目标是通过建立完善的数据，进而模拟疫情动态，从而科学地指导防疫工作。

2.2. 疫情信息可视化

信息系统的可视化展示与分析能够直观诠释隐藏在数据背后的信息，使数据能够更加形象生动便于大众接受理解。疫情期间，人们通过疫情地图不仅能够查询各地的实时疫情数据，还能根据可视化图表和地图了解疫情的变化和分布状况。百度、腾讯和今日头条等通过对国家卫健委公布的疫情数据进行实时地搜集整理，提取出有用的疫情数据，再进行可视化操作，以图表形式将疫情相关数据展示出来，方便群众更清晰直观地了解疫情新动向。其中百度地图迁徙大数据平台让人印象深刻，如图 1 所示百度疫情地图以及“疫情小区”专题地图、“发热门诊地图”等满足了用户特殊时期的出行、就医需求，了解疫情的扩散程度。

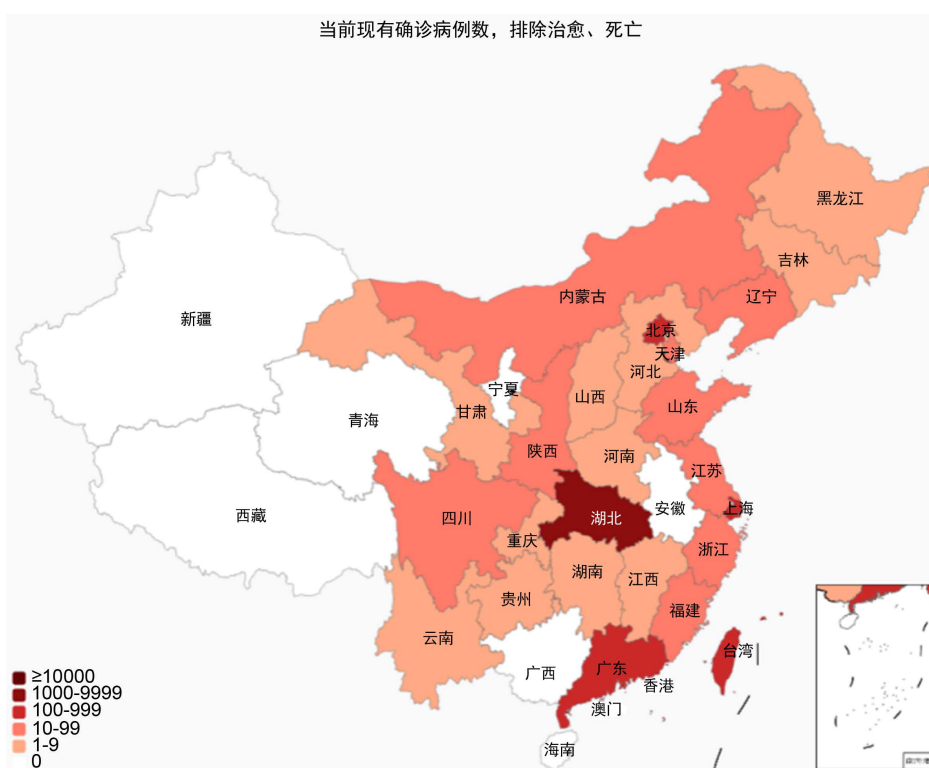


Figure 1. National distribution map of confirmed cases of Baidu

图 1. 百度疫情确诊病例全国分布地图

除此之外，如图 2 所示信息可视化还能将病毒更加生动形象地展示出来，详细分析新冠病毒的生成过程、传播途径、感染症状、病毒对比，第一部分为主体，用生物图例搭配文字说明形式，阐释感染过程、剖析病毒结构。第二部分分成三个小部分，分别介绍目前能够感染人类的七类冠状病毒，并按轻重

症状进行划分和病毒的传播过程和媒介以及感染冠状病毒后的临床症状。第三部分是数据对比，将历史上造成严重传染和伤亡的三种冠状病毒进行多维度的对比，清晰显示三者间的异同。使人们能更简单清晰地了解病毒的大致信息以及防控知识。

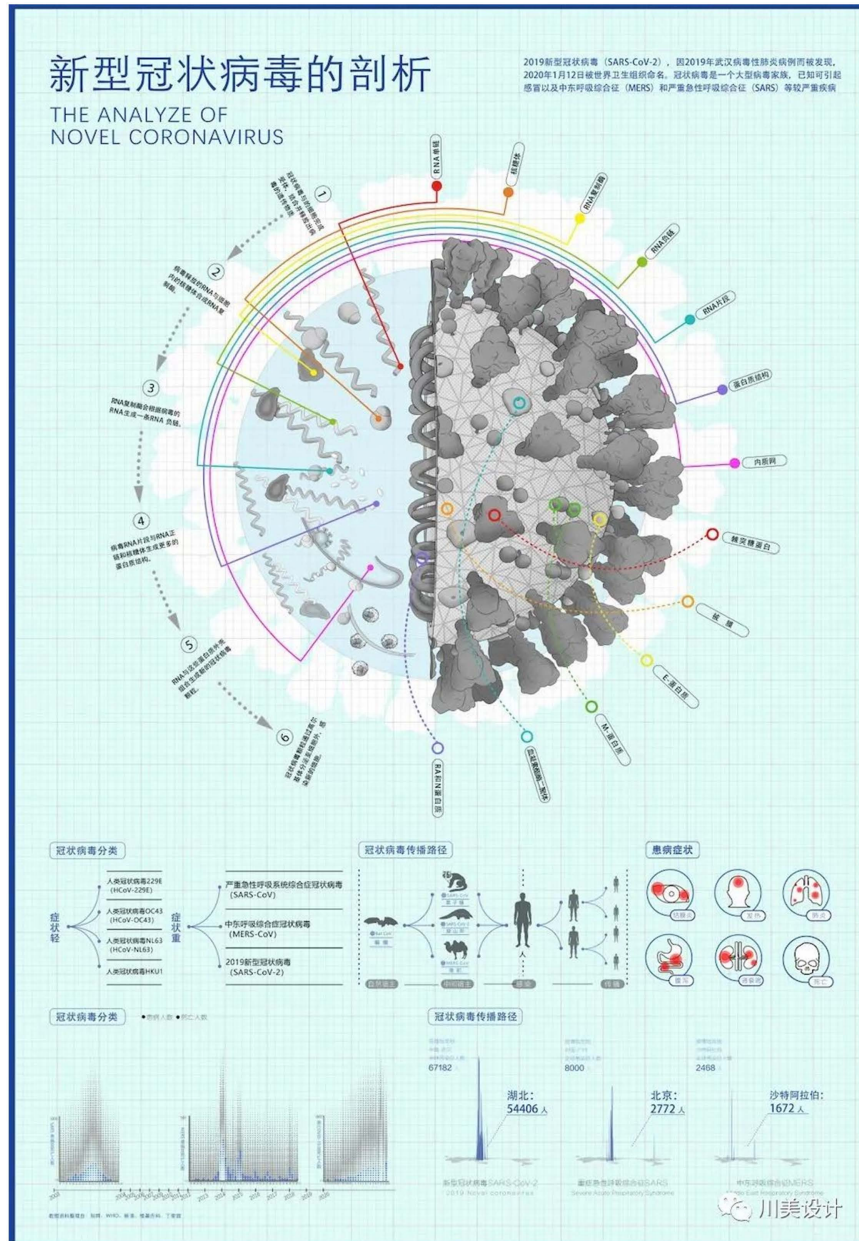


Figure 2. Virus analysis diagram of COVID-19
图 2. 新型冠状病毒剖析图

2.3. 疫情传播趋势预判

大数据时代下，以数治“疫”，时空大数据分析体系的建设对疫情防控疫情趋势预判带来重要影响，如何追踪感染者、密切接触者、疑似者都是至关重要的问题。大数据背景下信息系统的应用便能有效地解决这类问题，在这次新冠肺炎的疫情中，国家工信部通过直报系统搜集信息，分析了全国特别是湖北

和武汉地区的人口流动情况[1]。中国移动、联通、电信等中国三大运营商的人口的迁移数据，覆盖的范围广，连续性好[2]，都参与了工信部的大数据人口流动统计，每日按时上报送数据的统计报告，为相关政府部门进行轨迹追踪提供数据支撑。除了三大运营商之外，许多 App 通过内置的我定位系统，如阿里巴巴、百度和腾讯等互联网企业通过系统的功能来获取用户的地理位置信息[3]。通过这些手机信令数据对感染者进行追踪，对密切接触者进行挖掘，实现时空数据关联分析：快速描述海量空间数据结构，深刻洞察数据价值；构建多变量网络，实现精细化地理空间赋能；驻留区域排查，多维度洞察时空关系；多维要素连接，排查潜在感染小区。最后通过分析，获知相关信息，在疫情发生前期进行疫情传播趋势的预判，让各地部门疫情防控的效率得到了提升。信息时代，通过信息系统大数据的应用，为疫情传播趋势的预判提供了有力工具。

2.4. 大数据疫情模型

结合大数据技术，利用信息系统中的确诊、感染、治愈等数据，根据医学中相关的疫情传播规律能够建立动态的大数据疫情分析模型，为疫情防控提供技术支持。在本次疫情中，工信部通过搜集到的人口流动的数据集，采用大数据技术建立了疫情期间人口流动信息大数据分析模型，对人口流动趋势进行预判，从而分析出各省市的疫情传播趋势以及人员流动带来的影响，为疫情防控提前做好相应的准备。

联通“智慧足迹”根据确诊人群的活动数据建立了一个密切接触者的分析模型，该模型主要是根据已知感染者的数据情况进行分析，通过其活动范围来推测出风险区域，为疫情防控做出贡献[4]。此外，可通过该分析模型的计算来判断可能的密切接触者，预警可能发生疫情的新高风险区域，为社区级疫情防控工作提供技术支持。

“疫情树林”，结合流行病学信息，建立了一种非参数疫情时空模型，对疫情在不同防控措施下的发展进行情景模拟。从“疫情树林”得到疫情在历史上不同防控情形下表现的特点，从轨迹数据得到的人群流动在历史上不同防控情形下表现的特点，最后对未来疫情在不同防控情形下的发展进行预测，为防控措施的制订与实施提供科学依据。

疫情风险预测平台，根据大数据模型计算分析的结果对疫情危险进行等级划分，预测平台面向公众依据确诊人员区域为用户所在的小区设定风险等级，让人民群众能够预知风险区域，做好自身防控。

2.5. 疫情早期趋势预知

信息系统能通过已存储的信息来帮助医务工作人员捕获异常，提供更全面可用的数据信息，预知疫情发生的可能。信息系统能够让医务人员尽可能的早发现疫情趋势。武汉最早的不明原因肺炎患者是在12月8日通报的，该患者未出现传染的迹象没有得到关注。直到12月26日，医院的电子病历系统(EMR)中新存储了一条数据[5]，同时在当天上午 EMR 还存储了另外三条类似的数据，这几条类似的数据都包含了三个相同的关键词：华南海鲜市场、发热和咳嗽。通过类似数据分析提前推测疫情发生的可能性。

疫情中患者的完整数据档案，包括患者的病程记录、检验结果、治疗全过程等以电子病例的方式存储，能够让医务人员或者医生在给病人诊疗的同时获取到患者的完整数据，为护理临床决策等提供支持依据[6]。通过对电子病例系统的更好使用，搜集到的患者信息数据更加完善，发现病情传染的可能性也越大。就此次新冠肺炎疫情而言，若没有电子病例这类信息系统检测出相似数据，这次疫情的发现可能会更晚，对我们带来的危害也就越大。

2.6. 加快病毒检测诊断速度

信息系统分析技术对研究冠状病毒的结构、致病机理起到了重要的作用，从而为如何应对病毒，药物的研发以及疫苗的研究奠定基础。例如：基于序列比对的基因组信息分析能够得知病毒的特征结构和

与其相似的其他病毒，从而能够更快地了解病毒，为应对疫情病毒提供先决条件。由图 3 可知，通过几种冠状病毒的蛋白质长度序列解析比对[7]，加快病毒检测诊断速度和应对措施的研发。从全基因组结构上分析，通过基因组序列相似性的比较，找到相似的病毒。通过比较能够给研究人员提供重要信息：新型冠状病毒在致病机理、受体结合等方面可能与相似病毒有很多相似的点。这些通过信息分析出来的冠状病毒数据为以后的病毒防控，药物研发，疫苗研究等工作奠定首要基础。

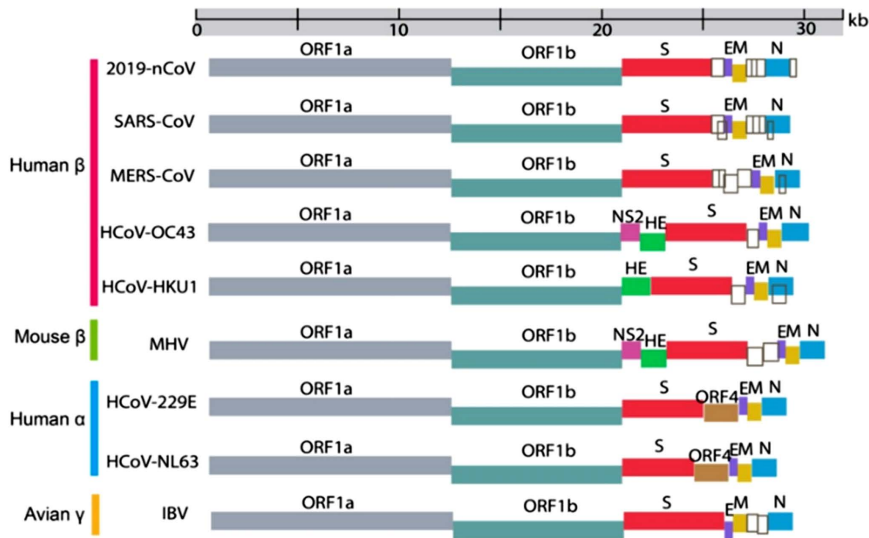


Figure 3. Sequence comparison of several coronavirus genomes
图 3. 几种冠状病毒基因组的序列比较

3. 存在的问题

3.1. 数据标准问题

信息系统中数据标准化是核心问题，数据标准化决定着信息资源的利用率与合理性[8]。数据标准化的忽视，会造成重复建设或者资源浪费等问题。目前来自不同厂商的电子健康档案系统没有实施标准化，各个厂商给医院提供的信息系统，有很多自己的文件、文档、数据格式，不仅医院之间的数据不能共享，而且医院不同科室的数据也难以实现集成。此次新冠肺炎疫情的过程中，数据变化与真实情况的差别过大，特别是在疫情前期和中期，数据标准存在问题。

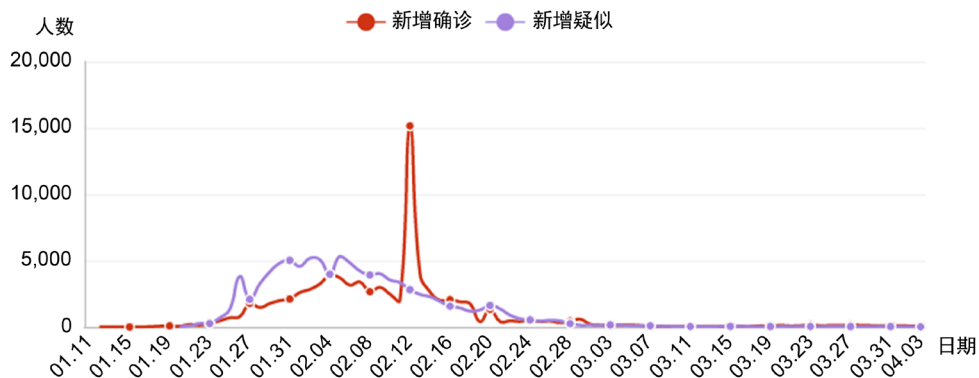


Figure 4. New diagnoses and new suspected trends across the country
图 4. 全国新增确诊、新增疑似趋势图

通过图 4 和初期的报告可以知道新冠肺炎确诊病例新增的趋势并不完全符合传染病的常规增长模型,其原因就是病例数据的判断标准问题。在这整个过程中,表面上是数据的波动问题,但实质上,数据波动的核心在于数据标准化问题、数据背后的标准波动。

3.2. 信息共享问题

信息系统的信息共享能够从信息的各方面出发加强信息的可利用率,充分利用信息的价值。病例数据仅仅是疫情大数据的组成部分,疫情防控还需要其他部门的相关数据作为分析和决策的支撑,例如人口数据、地理数据和交通数据等,因此数据共享不仅仅是医疗和公共卫生系统内部数据的共享,同时也需要与社会其他相关部门,如交通、公安、物资、运输等部门进行外部共享。如果信息共享存在问题,就会带来人力物力资源浪费,工作效率低下等等的问题。疫情防控期间湖北以及武汉各地的医疗资源需要信息比较零散没有形成很好的信息共享,这将会导致增援的物资分配出现分配不及时,物资积压等情况。捐助信息、需求信息的相互平衡,从捐助到使用的过程中,只有信息共享,物资的利用率才会更高[9]。

3.3. 信息分析处理问题

信息系统中的信息分析处理也是关键问题之一,传统的信息分析处理的数据主要是结构化的数据。然而目前随着信息化时代大数据的到来,半结构化以及非结构化的数据日趋增多,传统的信息分析处理已经很难胜任处理这些数据的任务。为了能够更好地让信息系统在医疗领域发挥作用,医疗健康非结构化的数据如基因图谱、医学影像、三维模型等半结构化和非结构化数据显得越来越重要[10],数据的多样性使得数据的呈现形式越来越复杂,而且数据也往往分布在不同的系统和平台上。数据的异构型、复杂性、多样性的问题是需要信息系统分析处理解决的,如电子病例的数据类型包括了文本、图像、图形和视频等多类型数据,存储着动态的多类型数据,这些数据的相关处理、实时分析都需要解决。

4. 对策与展望

针对数据标准问题,首先应该加强的是对于数据的有效管理,这就要求在国家层面统一开发标准和规范,各地医保信息化建设部门也要拥有一支高素质且稳定的专业技术队伍,在选择疫情医疗系统开发时才能够对信息系统项目按照标准化流程进行科学化、规范化管理,尤其注重掌握完备的技术文档资料,对系统的建设能严格按照标准实现自主可控。建立标准化的信息分类编码。各地严格按照国家标准化信息分类编码,即使各地确需扩充编码也应按照国家颁布的编码规则进行扩充,方能实现步调一致,互联互通。信息系统的质量与接口的标准化密切相关,接口标准化是信息标准化的一环。医疗保险信息系统与定点医疗机构、定点零售药店、银行、税务等信息系统进行数据交换时,必须执行制定的数据接口标准,包括数据项标准、数据格式、代码标准等,方能实现数据标准一致。

针对数据共享问题,在大数据时代,数据不再是部门的“专属品”,需要破除数据“部门私有”的旧观念,大力强调其“公共品”的属性。必须树立数据“共享增值”的理念。构建一体化、整合式的医疗系统大数据交换共享平台,通过数据共享,充分开发应用医疗大数据,产生新的价值,制定统一的地域医疗信息化标准,实现各级医疗机构平台兼容。在国家层面,我国即将建立起全国一体化的国家大数据中心,建设覆盖全国的国家大数据中心网络和政府数据开放平台,推进公共数据开放和基础数据资源跨部门、跨区域共享,国内一些地区在医疗信息共享方面已经有了很好的探索,例如由申康医院发展中心建设的上海市区域医疗信息共享模式,申康医院发展中心类似于公司运营的模式,由上海的 23 家医院组成。在实现信息共享的过程中,申康医院发展中心统一为这 23 家医院提供新的数据接口,与申康医院发展中心的系统进行连接,由此 23 家医院的所有信息都将自动流入申康医院发展中心,并由申康医院发展中心统一保存这些健康档案。

针对于信息分析处理的问题,采用数据整理的技术,进行前期数据解析与结构化处理、数据质量评估与数据清洗、数据集成和提纯等过程,使多源异构数据要更好地与其他数据集融合,比如本次疫情中的病例数据作为非结构化数据,需要结构化处理,对于中文电子病历进行词汇分割、语义提取等,影像数据需要基于医学的相关知识进行图像分析,将图像中的病灶转化为结构化的数据,包括数值数据和规范的文本数据。由于问题的复杂性,数据整理过程通常不是完全自动化的,而是需要用户介入的反复迭代和交互的过程,医疗数据的处理需要相关人员确认符合医疗知识的规则。

总的来讲,这次新冠疫情的防控是一场与时间赛跑、与病魔较量的斗争,也是对大数据在医疗信息系统中应用的实践证明。通过对疫情期间数据的实时采集、处理、分析、挖掘再输入医疗信息系统,加强疫情溯源和监测,能更为有效地促进信息透明,更好地维护社会大局稳定,大数据在疫情期间能够实时帮助发布权威信息,也能解决信息不对称的问题,有效压减谣言等虚假信息传播空间。同时普及科学防护知识,有效地满足了民众的信息需求,精准掌握疫区人员信息,定向精准指导各类人群的风险识别,居家隔离,到院就诊,有效引导群众理性看待疫情,增强自我防范意识和防护能力,能够在有限的时间内为疫情防控战略提供了更加全面的、自动化的管理以及各种服务,加速实现了线上线下深度结合,为信息公开、疫情监测、民生保障创造有利条件,为探索推进疾控体系现代化积累经验。目前信息化时代飞速发展,国家和国内各医疗机构也在加强信息化建设,信息系统在医疗领域的发展会越来越快。现在蓬勃发展的大数据技术将会给信息系统带来技术上的升级。总之,信息系统在医疗领域仍有很长的路要走,未来的信息系统还需加强建设发展。

参考文献

- [1] 罗名海,谭波,秦思娴,等. 大数据视角下的武汉市人口格局探析[J]. 地理空间信息, 2019(10): 1-6.
- [2] 赵时亮,高扬. 基于移动通信的人口流动信息大数据分析方法与应用[J]. 南京人口管理干部学院学报, 2014, 30(3): 20-26.
- [3] 蒋小荣,汪胜兰. 中国地级以上城市人口流动网络研究——基于百度迁徙大数据的分析[J]. 中国人口科学, 2017(2): 35-46.
- [4] 宁利君. 人员密集场所人群聚集风险监测与预警系统研究[J]. 安全, 2013(11): 59-60.
- [5] 叶全富,舒婷. 基于电子病历系统的医疗质量评价现状与趋势探讨[J]. 中华医院管理杂志, 2018, 34(7): 560-563.
- [6] 饶春梅,杨银玲,马旭梅,等. 基于结构化电子病历的护理临床决策支持系统的设计与应用[J]. 中国实用护理杂志, 2019, 35(1): 65-69.
- [7] Zhou, Y., Hou, Y., Shen, J., *et al.* (2020) Network-Based Drug Repurposing for Human Coronavirus. medRxiv.
- [8] 黄明科. 浅论信息系统数据标准化问题及对策[J]. 大众科技, 2011(6): 42-43.
- [9] 刘亚静. 网络信息化背景下医疗档案信息共享存在的问题及对策探讨[J]. 河北医学, 2014(5): 878-880.
- [10] 赵俊,孙亚丹. 基于 Hadoop 的医疗健康非结构化大数据分析研究[J]. 科技视界, 2017(36): 1-2, 74.