

基于大数据与 5G 技术的传染病智能监测预警处置系统设计与应用*

赵 坚 徐小卫 杨亚洲 虞 莹 赵 玉

(南通市疾病预防控制中心 南通 226007)

(南通市第一人民医院 南通 226001)

〔摘要〕 介绍传染病监测预警研究与应用现状, 详细阐述基于大数据与 5G 技术的传染病智能监测预警处置系统架构设计、多元异构数据采集和治理、模型设计和功能设计, 分析系统应用效果。试运行显示, 该系统有助于提升传染病监测预警水平和疫情处置效率, 实现全闭环管理。

〔关键词〕 大数据; 5G; 智能预警; 传染病监测; 传染病预警处置; 流行病学调查

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2023.05.003

Design and Application of an Intelligent Infectious Disease Monitoring, Early Warning, and Disposal System Based on Big Data and 5G Technology ZHAO Jian, XU Xiaowei, YANG Yazhou, YU Ying, Nantong Center for Disease Control and Prevention, Nantong 226007, China; ZHAO Yu, Nantong First People's Hospital, Nantong 226001, China

〔Abstract〕 The paper introduces the research and application status of infectious disease monitoring and early warning, elaborates the architecture design of an intelligent infectious disease monitoring and early warning and disposal system based on big data and 5G technology, the collection and management of multiple heterogeneous data, model design and system function design, and analyzes the application effect of the system. During the trial operation, the system improves the level of infectious disease monitoring and early warning and the efficiency of epidemic disposal, and realizes full closed-loop management.

〔Keywords〕 big data; 5G; intelligent early warning; monitoring of infectious diseases; management of infectious disease alerts; epidemiological investigation

1 引言

传染病一般指由各种病原体引起的疾病, 此类疾病通常在人与人之间、动物之间或人与动物之间进行

传播^[1], 具有危害大、传播迅速、早期隐蔽性强等特点^[2]。近年来传染病防控工作不断推进, 但仍然存在一些问题, 如信息系统敏感度有待提高、各传染病监测主体之间存在信息壁垒、业务协同度低、现场调查处置手段效率不高等。因此通过大数据、

〔修回日期〕 2023-03-29

〔作者简介〕 赵坚, 高级工程师, 发表论文 8 篇; 通信作者: 虞莹, 副主任技师, 发表论文 10 篇。

〔基金项目〕 工业和信息化部 5G+医疗健康应用试点项目“5G+5D‘五位一体’区域医联体重病云病房远程联合救治系统”; 江苏省南通市基础科学研究计划和社会民生科技计划项目“基于大数据和 5G 的南通市多点触发传染病预警综合平台研发与应用示范”(项目编号: MS22021027); 江苏省卫生健康委员会医学科研项目“人工智能和 5G 赋能不明原因传染病综合预警平台关键技术研究”(项目编号: Z2022094)。

5G、人工智能等新兴信息技术实现传染病实时监测、早期预警、精准处置在区域化传染病防控中具有重要意义^[3-4]。

2 现状分析

近年来信息技术高速发展，相关学者利用大数据、人工智能等技术进行了多种传染病监测预警方法实践^[5-6]。目前，国家疾病预防控制中心在全国范围使用传染病疫情和突发公共卫生事件网络直报系统进行监测数据的采集、自动分析、时空聚集性实时识别、预警信号发送，实现对法定传染病的早期自动预警，以及对全国各地传染病监测报告数据的分析与利用。该系统在法定传染病监测预警方面取得一定效果，数据仅来自全国各级医疗卫生机构直报，且在病例已确诊属于 39 种法定传染病之后才会上报，针对新发传染病监测预警的敏感性和时效性有待提高^[7]。因此建设一套标准化全闭环的区域化传染病智能预警处置系统作为补充显得尤为重要^[8]。

3 系统架构设计

经过对系统的功能、性能、安全性等进行全方位评估，南通市传染病智能预警处置系统最终以公共卫生数据中心为基础，采用浏览器/服务器方式搭建，通过政务外网互联，既满足系统建设需求，又在保证数据安全的同时具备一定可扩展性。

3.1 整体架构

系统整体架构包括安全基础层、设备层、数据层、应用支撑层和应用系统层 5 层，见图 1。其中安全基础层通过通信加密、身份认证、安全审计等底层安全策略确保系统数据环境可信安全。设备层从底层硬件出发为系统搭建可靠的运行环境。数据层为系统提供数据基础并对采集数据进行治理及质控。应用支撑层主要为应用系统提供决策支持，其中预警模型主要为传染病监测预警应用提供支持；身份识别信息、5G 通信模组、视频监控数据等主要为突发公共卫生事件的研判和处置提供支持。应用系统层为该系统提供监测预警、应急响应、智慧流调等主要功能。

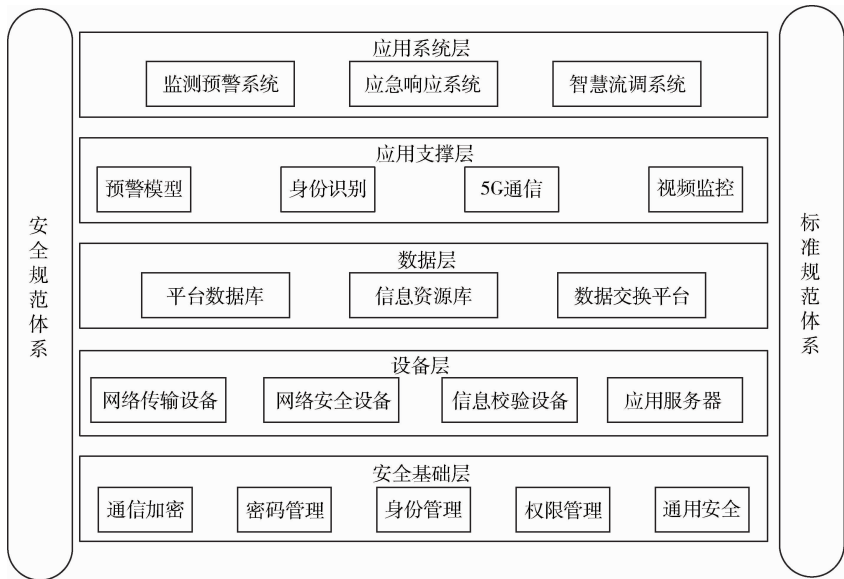


图 1 传染病智能预警处置系统架构

3.2 系统应用

基于上述架构，系统可以采用热力图、动态指

标、预警消息等可视化展示形式，结合地理信息系统 (geographic information system, GIS)^[9-10] 监控公共卫生事件风险因素发展情况。同时，通过构建风

险预测预警模型进行区域化传染病监测预警，在接收到预警信号后根据不同内容和风险指标详细研判分析，实现风险精准定位和全流程闭环处置。南通市公卫数据中心内还配备高速服务器集群等高性能计算资源，在移动终端设备中采用 5G 通信技术，进一步提升系统的可靠性和安全性。

4 多元异构数据采集和治理

4.1 数据来源

本系统主要采集电子病历、传染病直报系统、公共卫生、交通出行、学校缺课、重点场所和重点货物及食品和药品监测、环境监测等多渠道数据资源。采集数据主要来自南通市全民健康平台、国家疾病预防控制中心信息系统、南通市食品药品监督管理局系统、南通市出入境人员管理系统等。系统自 2021 年试运行以来，已采集各类数据 3.12 亿条，重点监测全市 47 类重点人群，共计 180 余万人，以及学校、医院等 400 余个重点场所。

4.2 数据采集治理流程

针对不同数据来源，执行国卫办医函〔2018〕1079 号《关于印发电子病历系统应用水平分级评价管理办法（试行）及评价标准（试行）的通知》、苏卫办医〔2011〕92 号《江苏省医院电子病历系统评价标准与细则（试行）》等卫生健康数据标准规范，确保系统兼容性^[11]。建立规范统一的主数据索引，运用 Hash 函数对原始数据进行脱敏，从数据的完整性、一致性、规范性、逻辑性等方面分析质量评估需求并制定评估规则，对采集数据进行质量分析与控制。基于 Hadoop 技术架构对系统采集到的数据进行分布式处理，为半结构化及非结构化临床数据提供分布式数据存储资源，并为自然语言处理模型算法提供分布式计算资源。对多源异构数据进行汇聚，利用大数据、自然语言处理技术对数据进行结构化、标准化治理，形成高可用的传染病监测预警数据库^[12]，见图 2。

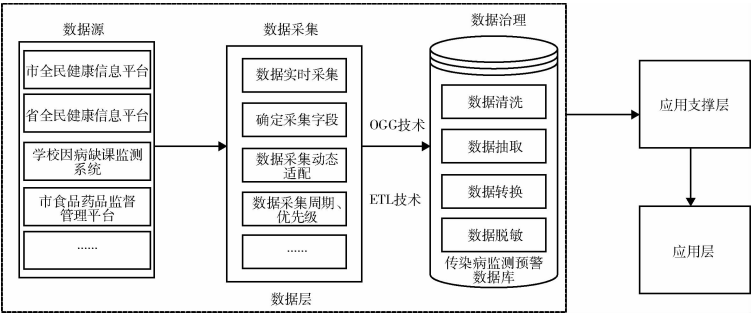


图 2 多元异构数据采集治理流程

5 模型设计

首先综合多种算法，解析传染病时空聚集特征；其次通过回顾性研究，分析传染病传播的风险因子；再次构建一个模型超市，采用仲裁组合策略加权整合候选模型；最后通过测试和自适应学习，进一步提升模型性能^[13]。目前该模型已经实现对肺结核、流行性感冒、手足口病、流行性腮腺炎等法定传染病的监测预警，并在试运行中取得预期效果。

5.1 综合多种算法，解析传染病时空聚集特征

采用时间聚集模式挖掘方法如聚类分析、集中度和圆分布法等^[14]，分析突发传染病的时间聚集性分布特征，包括年际发病特征和季节性发病特征，以实现预警前移；采用空间聚集模式挖掘方法如核密度估计和最近邻指数法等^[15]，分析突发传染病的空间聚集性分布特征；采用时空聚集模式挖掘方法如层级聚类法、时空扫描等^[16]，分析突发传染病的时空聚集性分布特征。

5.2 分析并确定影响传染病传播的风险因子

采用贝叶斯时空模型从时间和时空层面探讨传染病发病影响因素：假设区域 i 在时间 t 的某疾病发病或死亡人数为 y_{it} ，当发病或死亡率较低时，通常认为 y_{it} 服从泊松分布；当发病或死亡率较高时，则认为 y_{it} 服从二项分布，那么：

$$E(y_{it}) = e_{it} \theta_{it} \quad (1)$$

其中 E 代表期望值， e_{it} 是 i 区域在 t 时间的疾病期望发病或死亡人数； θ_{it} 是 i 区域在 t 时间疾病实际发病或死亡人数与期望发病或死亡人数的比值，也就是疾病发病或死亡的相对危险程度。连接函数采用 θ_{it} 的 $\log$ 函数，则贝叶斯时空模型公式如下：

$$\log(\theta_{it}) = \alpha_0 + \beta X_{it} + C \quad (2)$$

其中 α_0 为截距， β 为相关因素的回归系数， X_{it} 为 i 区域在 t 时间的相关因素。 C 为拟合的时空效应，可为互相独立效应或交互效应。贝叶斯时空模型的计算可以基于马尔科夫链蒙特卡罗算法，模型拟合优度可采用离差信息准则，该值越小表示模型拟合效果越好。

5.3 构建模型超市，采用仲裁组合策略加权整合候选模型

基于模型超市思想，针对不同传染病在不同应用场景下构建多组模型，采用组合策略根据误差大小进行加权组合，对不同阶段个体环境暴露水平进行估计，基于仲裁组合策略的加权思想进行模型整合，见图 3。

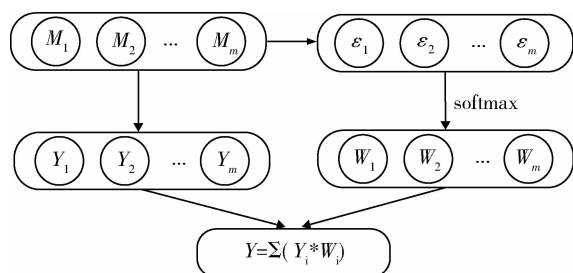


图 3 基于仲裁组合策略的加权整合设计

其中， M_i 表示各候选模型； Y_i 为各候选模型的预测值， $i \in \{1, \dots, m\}$ ； m 为候选模型个数，在不同环境暴露因素中可进行删减； ε_i 为各候选模型预

测误差； W_i 为经激活函数 softmax 计算输出的权重，取值为 $0 \sim 1$ 。首先，对各类基础模型 M_i 进行离线训练，即针对不同传染病在不同场景下训练多种不同模型，作为最终预测结果 Y_i 的一组基础学习器。训练后，返回训练好的所有模型集合 M ，在训练过程中，各模型不只输出其预测值 Y_i ，也持续输出各候选模型预测误差 ε_i 。误差度量针对不同场景的选择将有所不同，常用的选择是平均绝对百分比误差，因为此度量是相对度量，并且将误差归一化为百分比值。具体表达形式为：

$$\varepsilon = \sum_{i=1}^m \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times \frac{1}{m} \quad (3)$$

当数据量较少时，常用均方根误差作为损失函数，即绝对误差，并且保留实际值的大小。具体计算方法为：

$$\varepsilon = \frac{1}{m} \sum_{i=1}^m (Y_i - \hat{Y}_i)^2 \quad (4)$$

加权策略：加入 softmax 以获得各模型权重。softmax 将这些误差估计值的负数（对应的权重将更小）作为输入并返回概率分布。softmax 生成的各模型权重之和为 1，其具体形式如下：

$$\text{softmax}(\varepsilon_i) = \frac{e^{-\varepsilon_i}}{\sum_{n=1}^m e^{-\varepsilon_n}} \quad (5)$$

通过 softmax 函数可以将多分类的输出值转换为范围在 $[0, 1]$ 和为 1 的概率分布，即各模型 i 的权重 W_i 。最终，在仲裁组合策略生成各模型权重后，对各模型 M_i 输出结果进行加权，获得整合后的模型输出 Y ：

$$Y = \sum_{i=1}^m Y_i W_i \quad (6)$$

5.4 模型预警规则设计

时空预警是以南通市各区县为时空对象，针对特定传染病，取该病种在南通市各区县过去 5 年同一时期的计数组成集合 S_i ，用当前模型预测的病列数 Y_i 对比 S_i 的均值 K 及标准差 σ ，判定当前区域内发生传染病风险：当 $Y_i > K + 2\sigma$ 时判定为高风险，产生红色预警信号；当 $K + 2\sigma \geq Y_i \geq K + \sigma$ 时判定为中风险，产生橙色预警信号；当 $K + \sigma > Y_i \geq K$

时判定为低风险,产生黄色预警信号。

固定阈值预警是指系统通过实时采集数据,当达到一定阈值时发出警报,再通过模型自适应功能对阈值进行动态优化调整,从而提高预警的准确性和及时性。

5.5 模型性能描述、优化与调整

以南通市 2017—2022 年间传染病数据为测试集对模型进行测试评估,模型预测准确率超过 70%,生成预警信号的有效时间窗口覆盖率超过 80%,灵敏度超过 98%,特异度超过 70%。以 2021 年某小学一起手足口病事件为例,模型预测自 7 月 22 日起发展趋势,预计 8 月 23 日前后日新增确诊数小于 1,病例预测准确率为 79.5%。

为了进一步提升模型性能,通过对模型日常预测、预警工作的反馈(如漏报、误报等),采用自适应梯度算法、最陡下降算法等更新模型参数。定期对传染病进行时空聚集模型挖掘分析,利用最新数据定量更新传染病影响因素,优化、改良模型结构和参数,不断提高预测、预警的敏感性、准确性和及时性。

6 系统功能设计

6.1 传染病监测预警和处置

南通市卫生应急指挥中心通过大屏对南通市传染病情况进行实时监测,在收到系统发出的预警信号后,南通市卫生应急指挥中心值守工作人员立即核实,并通过区域化传染病联防联控机制上报,依据预警事件风险等级启动应急处置流程,由卫生防疫部门立即派出流调小组进行流行病学现场调查和溯源工作。系统通过连接现场的移动 5G 视频终端将高清视频信号实时传输至南通市卫生应急指挥中心大屏,领导和专家可通过观看大屏对现场情况进行实时掌控和指挥调度。

6.2 基于 5G 技术的应急响应处置

系统基于 5G 网络高速率、低时延的特性,将流调现场数据和南通市卫生应急指挥中心无缝对接,利

用 5G 单兵设备内置音频软件,将现场录音识别为文字,经过简单人工确认和格式调整后可以即时传输至南通市卫生应急控制中心供指挥者参考,也可以纳入现场调查和处置报告中,有效减少手工记录工作量。5G 单兵设备还具备卫星定位功能,可以通过 GIS 直观显示该起突发公共卫生事件所处的位置环境、应急处置队伍分布等情况,并通过 5G 单兵设备将现场情况实时传输至南通市卫生应急指挥中心,便于领导和专家掌握事态发展和制定方案。

6.3 智慧流调和处置

系统通过命名实体识别,结合基于规则的正则表达式技术和基于大规模语料深度学习的模型,采用结构化规则和深度学习两种技术路线,通过对特征的提取和识别智能生成流调报告,解决传统流调报告效率低,需要多人、多次进行信息补充及完善的问题^[17-19]。

7 系统应用效果

7.1 传染病智能预警和处置应用

系统自 2021 年 6 月起试运行,预计于 2023 年 6 月正式上线。截至 2023 年 2 月底,系统已采集有效数据 3.12 亿条,处置各类预警信号 52 000 多个,根据南通市委市政府相关文件精神,目前模型已将出入境人员、发热门诊患者等 47 类传染病重点监测人群共约 180 万人及医院、学校、机场、车站等 400 余处重点监测场所纳入重点监测范围。以重点场所监测预警为例,系统通过多算法解析传染病时空聚集特征分析风险时段和风险区域,使用贝叶斯时空模型分析风险传播因子,采用加权整合生成的预警模型进行实时监测预警。在 2021 年 6 月 1 日—9 月 30 日期间南通市共产生学校热点预警信号 221 次,涉及病例 698 例,已核实流感病例 512 例、手足口病例 118 例、肺结核病例 2 例,其中 4 起预警较国家传染病自动预警提前 4~7 天,实际发生 3 起,基本达到预期目标。

7.2 应急处置和智慧流调应用

截至 2023 年 2 月底已处置各类突发公共卫生事

件 30 余起, 智能生成流调报告 2 700 余条, 确保各环节时间节点有据可查, 不但有利于事后复盘, 还能够进一步优化工作流程, 极大地提升疫情应急处置效率。

8 结语

本系统采用安全基础层、设备层、数据层、应用支撑层和应用系统层 5 层架构, 通过整合医疗、疾病预防控制等多元异构数据资源, 结合多算法和贝叶斯时空模型建立智能监测预警模型, 实现风险预警关口前移, 提升传染病早期监测预警能力, 实现监测预警事件的全流程跟踪及处置。针对疫情处置方面存在的问题, 建设数字化智能流调系统, 提升疫情应急处置效率, 达到早研判、早处置、早阻断目标。该研究方法也可供其他同类地区、单位参考, 作为进一步优化区域传染病防控的管理方法和手段。

在系统建设过程中遇到部分机构提供的的数据质量不高、数据接口对接难度大等问题, 通过加大政府行政推动力度、优化数据采集技术、增加数据采集频率、调整数据采集流程等方法加以解决。在系统正式上线后将从实际需求出发, 继续拓展传染病监测种类和监测数据来源, 结合新技术对系统进行定期升级维护, 进一步提升系统使用效能和实用性。

参考文献

- 1 鲁芳芳, 胡池, 徐承中, 等. 区域多点触发传染病预警机制的建立与应用 [J]. 中国卫生信息管理杂志, 2022, 19 (6): 798 - 803.
- 2 蒋天宏, 周正元, 龚利强, 等. 基于大数据的传染病监测预警处置系统设计与分析 [J]. 中国数字医学, 2022, 17 (11): 116 - 120.
- 3 GINSBERG J, MOHEBBI M H, PATEL R S, et al. Detecting influenza epidemics using search engine query data [J]. *Nature*, 2009, 457 (7232): 1012 - 1014.
- 4 张众, 王忠岳. 传染病早期预警机制建设与对策建议 [J]. 医院管理论坛, 2021, 38 (10): 12 - 15.
- 5 李晓洁, 丛亚丽. 从“谷歌流感趋势”预测谈健康医疗大数据伦理 [J]. 医学与哲学, 2019, 40 (14): 5 - 8.
- 6 ODLUM M, YOON S. What can we learn about the Ebola outbreak from tweets [J]. *American journal of infection control*, 2015, 43 (6): 563 - 571.
- 7 杨世宏, 肖丹, 姚伟, 等. 大连市传染病自动预警系统构建及效果评价 [J]. 中国公共卫生, 2015, 319 (11): 1505 - 1508.
- 8 来晓强, 殷小娟, 张彦青, 等. 定西市传染病自动预警系统运行效果分析 [J]. 疾病预防控制通报, 2016, 31 (4): 22 - 24.
- 9 WEI G, TAO X, KENGMING T. M - estimator based on-line sequential extreme learning machine for predicting Chat-IC time series with outliers [J]. *Neural computing and application*, 2017, 28 (12): 4093 - 4110.
- 10 LI Q, GUAN X, WU P, et al. Early transmission dynamics in Wuhan, China of novel coronavirus - infected pneumonia [J]. *The New England journal of medicine*, 2020, 382 (13): 1199 - 1207.
- 11 陈冰冰, 李保军. 海曙区学校传染病聚集性疫情监测分析 [J]. 预防医学, 2018, 30 (8): 812 - 815.
- 12 吴亚琼, 韩鹤, 王雅婧. 基于 SEIR 传染病模型的应急救援物流调度优化 [J]. 江苏科技大学学报 (自然科学版), 2019, 33 (1): 87 - 92.
- 13 王梦莹, 张文丽, 高玥, 等. 基于大数据的医院真实场景传染病预警系统研究 [J]. 中国医院管理, 2022, 42 (3): 1 - 5.
- 14 杨维中, 兰亚佳, 李中杰. 传染病预警研究回顾与展望 [J]. 中华预防医学杂志, 2018, 48 (4): 244 - 247.
- 15 OUYANG L W, YUAN Y, WANG F Y. A blockchain - based framework for collaborative production in distributed and social manufacturing [C]. Piscataway: 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), 2019.
- 16 SALATHE M. Digital pharmacovigilance and disease surveillance: combining traditional and big - data systems for better public health [J]. *Journal of infectious diseases*, 2016, 214 (S4): 399 - 403.
- 17 杨雯, 崔冬. 我国传染病风险预警制度所面临的风险挑战与因应建议 [J]. 医学与法学, 2022, 14 (1): 59 - 66.
- 18 侯攀锋, 何文英. 传染病监测预警现状研究 [J]. 医学信息, 2021, 34 (18): 36 - 38.
- 19 陈媛, 丘美娇, 陈宝, 等. 综合医院发热门诊信息化建设对传染病监测预警及登记报告的影响 [J]. 海南医学, 2021, 32 (17): 2293 - 2297.