文章类型：论文 | 刊号（ISSN）：2972-4457(P) / 2972-452X(O)

统计建模与数据挖掘在公共卫生监测领域的实践与展望

吴祺
北京信息科技大学
DOI:10.32629/ffcr.v3i11.22354

[摘　要] 公共卫生监测是疾病防控、卫生资源调配的基础，传统的监测模式存在着数据利用率低、预警滞后、风险识别片面等问题。统计建模与数据挖掘技术融合应用，可以挖掘多源卫生数据的价值，实现疾病流行规律分析、风险精准预警、卫生态势动态评估。本文梳理两类技术在传染病监测、慢性病防控、公共卫生风险预警中的实际应用场景，分析目前技术落地过程中存在的问题，结合行业发展趋势，提出技术优化和落地升级的方向，为公共卫生监测智能化、精准化发展提供参考。
[关键词] 统计建模；数据挖掘；公共卫生监测；疾病预警；精准防控
中图分类号：R195.4　文献标识码：A

Practice and Prospect of Statistical Modeling and Data Mining in Public Health Surveillance

Qi Wu
Beijing Information Science and Technology University

Abstract: Public health surveillance is the foundation of disease prevention and control and health resource allocation. Traditional surveillance models have drawbacks including low data utilization, delayed early warning and one-sided risk identification. The integrated application of statistical modeling and data mining can tap the value of multi-source health data to analyze epidemic patterns, realize precise risk early warning and dynamically assess public health status. This paper reviews practical application scenarios of the two technologies in infectious disease surveillance, chronic disease prevention and control and public health risk early warning, analyzes the existing problems in technology implementation, and proposes directions for technical optimization and application upgrading combined with industry trends, so as to provide references for the intelligent and precise development of public health surveillance.
Keywords: statistical modeling; data mining; public health surveillance; disease early warning; precise prevention and control

引言

随着公共卫生体系数字化转型的推进，医疗就诊数据、环境监测数据、人口健康数据、舆情态势数据等多源数据体量不断增大，传统依靠人工统计、常规报表的监测模式已经不能满足复杂多变的公共卫生防控需求。公共卫生监测的核心需求已经从单纯的病例统计、数据汇总转变为趋势预判、风险溯源、精准施策[1]。统计建模依靠严谨的数理逻辑量化卫生事件的发展规律，数据挖掘技术可以突破传统数据分析的局限，挖掘海量数据中的隐性联系和潜在风险[2]。二者结合可以很好地弥补传统监测的不足，广泛应用于各种公共卫生场景，成为现代公共卫生防控的主要技术支撑。本文根据近年实践成果，对其应用现状、存在的问题以及发展前景进行系统总结。

1 统计建模与数据挖掘在公共卫生监测中的实践应用

1.1 传染病动态监测与流行预测

传染病监测是两类技术应用最成熟的应用场景。传统 SIR、SEIR 等经典统计传播模型可以量化易感人群、感染人群、康复人群的动态变化，模拟传染病传播趋势，适合于流感、新冠病毒等呼吸道传染病的常规流行态势分析，可以为阶段性防控策略调整提供数据支持。随着多源数据融合需求的提高，数据挖掘技术也逐渐实现模型升级，利用随机森林、时序挖掘、深度学习等算法把气象、人口流动、区域防控措施等混杂变量整合起来，解决传统模型变量单一、预测偏差大的问题[3]。目前主流的智能监测系统可以利用社交媒体、医疗机构就诊数据、污水监测数据等多渠道的信息，快速捕捉到疫情早期的异常信号，实现疫情的超前预警和短期精准预测，大幅缩短了传统监测的滞后周期[4]。

Copyright © This work is licensed under a Commons Attribution-Non Commercial 4.0 International License.

1.2 慢性病风险筛查与趋势研判

慢性病防控中两类技术主要用在群体风险评价和发病趋势分析上。利用居民健康体检、慢病随访、医保就诊等海量数据，用统计回归模型量化年龄、作息、基础病史、环境因素与高血压、糖尿病等慢性病发病的关联强度，找出区域慢病高发危险因素[5]。同时使用数据挖掘聚类算法可以准确地把高危人群、普通人群分成不同的层次，找到区域慢病高发聚集区，改变传统的粗放式筛查方式。基层公共卫生机构根据分析结果，对高危人群开展干预、健康宣教等精准服务，有效地降低了区域慢性病发病增长率，提高了慢病管理效率[6]。

1.3 公共卫生突发风险智能预警

突发公共卫生事件具有突发性、扩散性强的特点，对监测的时效性、精准性要求很高。目前行业普遍采用统计时序建模加文本数据挖掘的融合模式搭建自动化预警体系[7]。利用自然语言处理技术抓取全网卫生相关舆情、异常就诊数据、公共环境异常数据，使用时空统计模型分析数据波动规律，识别非常规卫生风险信号，可以快速识别食物中毒、群体性不明原因疾病、区域性污染致病等突发风险，实现从“事后处置”向“事前预警、事中管控”的模式转变，有效提升公共卫生应急处置能力[8]。

2 当前应用存在的问题

现阶段统计建模和数据挖掘技术在公共卫生监测中仍存在很多现实不足。数据质量参差不齐，多源卫生数据格式不统一、信息缺失、重复冗余，不同区域、不同机构数据壁垒还没有完全打破，造成模型分析精度受限。二是模型适配性差，大部分高精度机器学习模型可解释性差，不能满足公共卫生防控的合规性、实操性要求，传统统计模型智能化程度低，不能适配海量复杂数据处理的需求。三是基层落地能力欠缺，先进的模型和算法大多被省级、市级的公共卫生平台所使用，基层机构缺少专业的技术人员和算力支持，技术应用存在明显的层级差别。

3 发展展望

随着大数据、人工智能技术的不断更新以及公共卫生防控精细化需求的持续提高，统计建模和数据挖掘在公共卫生监测领域的应用将向着精准化、智能化、普惠化、规范化全方位升级，突破目前技术落地的瓶颈，全面赋能现代化公共卫生体系建设，主要发展方向有四个方面。

第一，推进统计和机器学习融合模型的研发和应用，提高场景适应性。目前单一统计模型算力不足、智能算法可解释性差的痛点，严重阻碍了技术在公共卫生监测中的官方应用和推广。未来行业将重点研发统计建模和数据挖掘相结合的混合模型，依靠传统统计模型的数理优势保证分析结果的科学性、可解释性，结合深度学习、集成学习等数据挖掘算法提高海量数据的特征挖掘和预测精度。根据传染病传播、慢病风险评估、突发卫生事件预警等不同场景，定制化优化模型参数，解决单一模型适配性差、预测误差大的问题，兼顾技术精准度和行业实操合规性[9]。

第二，健全公共卫生大数据治理体系，筑牢技术应用基础。数据质量是模型分析和智能监测的前提，未来会不断打破医疗、疾控、环境、人社等各个部门的数据壁垒，创建统一标准化的公共卫生数据共享平台。统一各类卫生监测数据的录入格式、统计标准、存储规范，建立常态化的数据清洗、纠错、脱敏、更新机制，有效解决数据缺失、冗余、失真、不互通等问题。完善数据安全管理制度，在保证数据安全、保护居民隐私的基础上，最大限度地挖掘多源数据的联动价值，为建模分析、风险挖掘、趋势研判提供高质量的数据支撑[10]。

第三，推进智能监测技术基层下沉，缩小区域应用差距。目前先进的数据建模和挖掘技术主要应用于省市一级的公共卫生机构，基层疾控中心、社区卫生服务中心技术算力不足、专业人才匮乏，造成监测工作仍然依靠传统的手工方式。未来将重点推进技术普惠化发展，开发轻量化、可视化、易操作的智能监测系统和分析工具，简化模型的操作流程，不需要专业的算法能力就可以完成数据统计、风险分析、趋势预警。同时常态化开展基层公共卫生技术培训，搭建线上线下相结合的指导体系，提高基层人员数据应用能力，实现省、市、县三级公共卫生智能监测体系全覆盖。

第四，扩大技术应用场景，创建全域智能监测体系。现阶段技术应用主要集中在传染病、慢性病核心领域，场景单一。未来会不断拓展技术应用的范围，把统计建模、数据挖掘技术应用到食源性疾病溯源、环境健康风险评价、妇幼健康监测、老年健康管理、职业卫生防控等各个领域。通过多维度数据联动分析，找出各类公共卫生风险的诱发因素和传播规律，创建全方位、动态化、立体化的公共卫生智能监测网络，实现各类卫生风险的早发现、早研判、早处置，全方位提升公共卫生防控的科学化、精细化水平。

4 结论

统计建模和数据挖掘技术相结合，彻底改变了传统的公共卫生监测方式，有效地提高了疾病监测、风险预警、健康管理的科学性，在传染病防控、慢性病管理、突发公共卫生事件处置等方面起到了重要的作用。目前技术应用还存在数据治理、模型适配、基层落地等各方面的限制，不能充分发挥数据赋能的价值。未来要不断改进模型算法、健全数据体系、推进技术普惠，促使智能监测技术同公共卫生业务深度融合，为创建现代化公共卫生防控体系赋予有力的技术支撑。在公共卫生数字化转型大背景之下，两类技术的创新应用会成为解决传统监测痛点、提高公共卫生治理能力的主要手段。

Copyright © This work is licensed under a Commons Attribution-Non Commercial 4.0 International License.

不断推进技术迭代和场景落地，可以有效地筑牢公共卫生安全防线，使居民的全民健康水平稳步提高。就长远而言，技术、数据同公共卫生业务深度耦合，会促使我国公共卫生监测工作步入全域智能、精准高效、全民普惠的新阶段。

[参考文献]

[1]郭占春.公共卫生监测对传染病预防控制的影响[J].中国卫生产业,2025,22(17):132-134,138.

[2]蒋舒仪.基于多源数据和人工智能模型的公共卫生监测系统[J].计算机应用文摘,2025,41(24):130-132.

[3]刘红慧,官旭华,童叶青,等.传染病多点多渠道监测预警体系建设探索——以湖北省为例[J].医学信息学杂志,2025,46(5):19-23,28.

[4]霍大柱,张婷,崔金朝,等.呼吸道传染病多源监测数据整合与预警指标应用研究进展[J].中华预防医学杂志,2025,59(8):1311-1319.

[5]张艳,王彦贺,林胜红,等.基于多数据融合的全球传染病监测预警信息系统的建立与初步应用[J].中华医院感染学杂志,2025,35(24):3768-3774.

[6]赵坚,徐小卫,杨亚洲,等.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[J].医学信息学杂志,2023,44(5):14-19.

[7]李丹彤,王梦莹,贾末,等.区块链数据共享监测与智能融合系统设计与应用[J].中国卫生信息管理杂志,2025,22(2):291-297.

[8]汤舒涵,戴智翔,王崴,等.人工智能在传染病监测预警中应用的现状与挑战[J].中华流行病学杂志,2025,46(10):1886-1891.

[9]刘民,梁万年,胡健,等.重大突发传染病智能化主动监测预警系统设计研究[J].中国工程科学,2024,26(6):65-76.

[10]马宇航,殷一,王凯,等.传染病多阶段、多场景监测预警多样化方法体系的研究与思考[J].中华预防医学杂志,2023,57(10):1529-1535.

作者简介：

吴祺（2005.10-），女，汉族，北京人，本科，研究方向为公共卫生。

Copyright © This work is licensed under a Commons Attribution-Non Commercial 4.0 International License.