

大数据在突发公共卫生事件预测和监测中的算法优化

翁心睿

北京化工大学，北京

摘要 | 突发公共卫生事件具有复杂性和不可预测性，这对传统监测体系构成了巨大挑战。文章提出“轮组”模型，借助大数据技术达成风险监测、信息采集、分析研判和预测支持，提升决策效率和精准性。针对当前数据应用中存在的虚假信息难辨、信息孤岛、隐私泄露及技术创新持续性不足等问题，提出优化方向，包括强化数据采集标准、完善数据资源体系、保护个人健康数据安全及建立长效技术创新机制等。此外，文章强调场景需求驱动下的大数据应用，可实现对多源数据的实时整合和分析，推动疫情预测、监测的智能化和自动化。通过总结实践经验与理论模型，文章为大数据赋能公共卫生事件监测提供了系统性思路，为风险治理和应急管理的创新提供了参考。

关键词 | 大数据；疫情监测；风险预测；轮组模型

Copyright © 2025 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>



一、引言

频发的突发公共卫生事件因其紧急性、复杂性和不可预测性，已成为全球

基金项目：2020年北京市社会科学基金青年项目“突发公共卫生事件信息公开法律规制的完善”（编号：20FXC029）的阶段性研究成果。

作者简介：翁心睿，北京化工大学法学与知识产权研究所助理研究员，研究方向：破产法、数据法。

文章引用：翁心睿. 大数据在突发公共卫生事件预测和监测中的算法优化 [J]. 法学进展, 2025, 7 (1): 93-104.

<https://doi.org/10.35534/al.0701009>

卫生体系和社会管理面临的巨大挑战^①。新冠肺炎疫情的暴发，暴露了卫生系统在应对未知病原体 and 大规模传播时的脆弱性，传统的监测和管理手段难以适应疫情暴发的速度和规模。因此，亟需更为创新和智能化的工具，以提升对公共卫生事件的实时感知和精准预测能力。随着新一轮科技革命和产业变革的深入推进，数据作为关键生产要素的价值日益凸显。充分发挥数据的乘数效应，创新应急管理方式，对优化突发公共卫生事件预测和监测具有重要意义。

二、突发公共卫生事件监测预测现状

突发公共卫生事件具有的突发性和不确定性，对预测和监测工作提出了更高要求，需要更为稳健、高效、安全的预测和监测方式。当前学界对于大数据驱动下的突发公共卫生事件的监测和预测已有大量研究，能够促进风险防控的精准性、有效性、快速性、可视性，提升突发公共卫生事件监测和预测的效率和质量。已有研究利用社交媒体网络中用户搜索或发布的言论等数据来获取公众面对、关注的卫生话题，如布罗尼亚托夫斯基（Broniatowski）等利用 Twitter 上用户发布的流感症状相关信息，以 85% 的准确率成功预测流感患病率^②。金斯伯格（Ginsberg）通过谷歌搜索查询，能够跟踪人群中的流感样疾病，准确估计美国每个地区每周的流感情况^③。我国的袁清玉教授首次利用百度的搜索查询数据来模拟、监测中国的流感活动，误差小于 11%^④。同时，社交网络上的信息也提供了个体间的时空关系网络，能够对疾病或者可能发生疾病的人群进行快速、集中的监测，并整合相关主体进行集中资源的统一治理^⑤。如在新冠肺炎疫情

① 范仓海，施思. 突发性公共卫生事件中的大数据治理：现实困境与路径优化——以新冠肺炎疫情防控为例 [J]. 卫生软科学, 2022, 36 (8) : 81-85, 96.

② Broniatowski D, Paul M, Dredze M. National and Local Influenza Surveillance through Twitter : An Analysis of the 2012-2013 Influenza Epidemic [J]. PLoS One, 2013, 8 (12) : e83672.

③ Ginsberg J, Mohebbi M, Patel R, et al. Detecting Influenza Epidemics Using Search Engine Query Data [J]. Nature, 2009, 457 (7232) : 1012-1014.

④ Yuan Q Y, Elaine O, Ben F L, et al. Monitoring Influenza Epidemics in China with Search Query from Baidu [J]. Chinese Influenza Epidemic, 2013, 8 (2) : 1-7.

⑤ 陈娟，郭雨丽. 社交媒体与疫情：对公共卫生事件的预测、沟通与干预 [J]. 新闻记者, 2020 (4) : 60-69.

情期间,上海市增设“一网统管”防疫专页,建立市、区、街镇三级城运中心,推进实时动态“观、管、防”一体化,为各级政府部门和机构提供可视化的决策辅助工具,实现“一屏观上海、一网管全城”^①。浙江省利用全省统一的公共数据平台“浙政钉”,建立省、市、县、乡、村五级的抗疫组织通讯录,节约一线的人力物力资源;通过“新冠智能诊断系统”辅助一线影像科医生快速完成疑似新冠患者的影像诊断;开发出“疫情防控驾驶舱”为各级政府领导和疫情防控工作人员提供随时随身的可视化决策辅助工具^②。

然而,近年来,大数据赋能的突发公共卫生事件预测和监测仍然存在一系列问题。一是在数据获取方面,信息虚实难辨。在新冠肺炎疫情期间,社交平台因缺少监督和约束,虚假信息泛滥,真假信息相互交织,多渠道的信息以及信息表达的模糊性增加了判定虚假信息的难度。二是在数据处理中,存在信息壁垒。公共卫生数据往往涉及多个政府部门、医疗机构、科研单位,以及私营企业,各自形成了信息孤岛^③,缺少有效的协作机制和充分信任,数据系统之间缺乏互联互通,从而导致数据共享存在障碍。三是信息保护存在漏洞,数据泄露风险较大。我国大数据平台在预测、监测突发公共卫生事件方面存在的局限性,凸显了亟需优化大数据驱动下的突发公共卫生事件预测与监测模型的紧迫性,从而最大限度地减少突发公共卫生事件所带来的风险。

目前,关于大数据在应急管理中的应用研究较为分散,缺乏对大数据赋能预测和监测突发公共卫生事件的深层运作逻辑和未来图景的研究^④。本文利用“滑轮组”模型细致拆分预测和监测的各个环节与大数据的联系,重点探讨大数据赋能突发公共卫生事件预测、监测的具体方式和内在逻辑,分析优化和提升突发公共卫生事件预测、监测数据平台的建设思路,提高突发公共卫生事件

① 张懿. 数据赋能,牵引带动超大城市治理现代化[N]. 文汇报,2022-06-16(1).

② 史晨,马亮. 互联网企业助推数字政府建设——基于健康码与“浙政钉”的案例研究[J]. 学习论坛,2020(8):50-55.

③ 邓源,任翔,黄硕,等. 大数据在传染病监测预警中的主要研究与应用进展[J]. 疾病监测,2022,37(8):1003-1009.

④ 温志强,付美佳. 大数据赋能政府应急管理模式变革:类型归结与未来向度[J]. 行政论坛,2023,30(4):79-86.

预测、监测能力。

三、大数据下突发公共卫生事件监测和预测的“轮组”模型

现有的大数据赋能突发公共卫生安全应对的相关动力学模型比较丰富。包括提炼公共卫生应急治理智慧性表征的“基础层——动力层——应用层——导向层”大数据驱动公共卫生智慧治理逻辑框架^①、解释技术应用和应急管理创新之间的相互作用关系的双螺旋赋能机理模型^②、结合突发事件发展阶段及不同阶段管理目标的“危险管理双周期曲线模型”^③、阐述大数据与应急管理双向聚合的大数据赋能公共卫生安全风险治理的“轮组式”模型^④。综合可以看出，当前对于大数据赋能突发公共卫生事件的模型主要聚焦在公共卫生事件发展的全过程，更多关注非常态化的公共卫生事件发生过程中利用大数据进行应急驱动，并没有聚焦在公共卫生事件发生前的数据赋能。

由于突发公共卫生事件如同生物体一样具有潜伏期、征兆期、发展期、衰退期、消亡期，笔者将研究重点集中在突发公共卫生事件处于潜伏期和征兆期的预测、监测任务，以公共卫生安全风险治理的“轮组式”模型为基础，构建突发公共卫生事件“定滑轮”模型和大数据下突发公共卫生事件监测和预测“滑轮组模型”，通过对比这两个模型，阐述大数据赋能下的突发公共卫生事件监测、预测能力显著提升，并分别拆分数据端和预测、监测端的各个环节更清晰地阐明大数据的赋能作用，同时对现有文献和理论进行归纳总结，丰富研究模型的理论性与可靠性。

① 陈潭，王鹏. 大数据驱动公共卫生应急治理的智慧表征与实践图景 [J]. 电子政务, 2021 (6): 85-99.

② 詹承豫，高叶，李治博. 应急管理中技术赋能的双螺旋机理研究——以人脸识别技术在疫情常态化防控中的应用为例 [J]. 行政管理改革, 2021 (7): 68-78.

③ 郭捷，杨立成，孙子旭. 基于科技视角与双周期模型的我国突发事件危机管理研究——以新型冠状病毒危机事件为例 [J]. 科技进步与对策, 2020, 37 (14): 8-13.

④ 宋劲松，夏霆. 大数据对公共卫生安全风险治理的赋能机理研究——以新冠肺炎疫情防控为例 [J]. 行政管理改革, 2022 (4): 21-29.

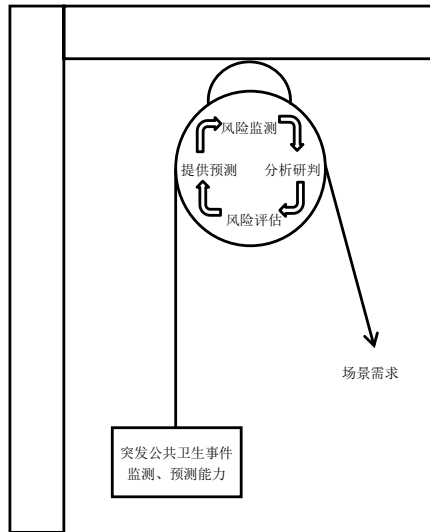


图1 突发公共卫生事件监测和预测的“定滑轮”模型

Figure 1 The “Fixed Pulley” model for monitoring and predicting public health emergencies

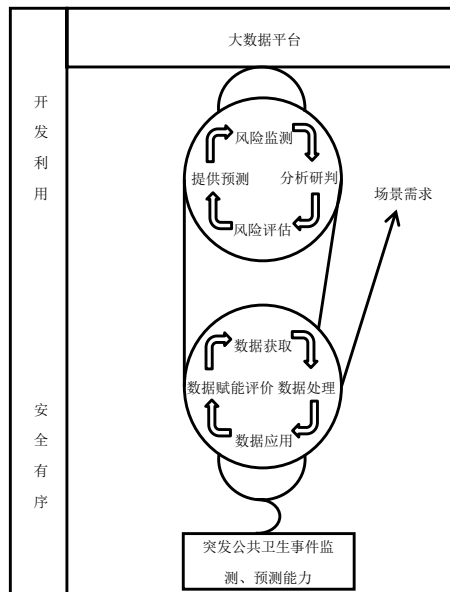


图2 大数据下突发公共卫生事件监测和预测的“滑轮组”模型

Figure 2 The “Compound Pulley” model for monitoring and predicting public health emergencies with big data

在突发公共卫生事件处于潜伏期和征兆期，面对场景需求，对风险的持续监测、预测如同“定滑轮”模型（图1）中的定滑轮，持续不断地进行循环流程。在“滑轮组”模型（图2）中，“定滑轮”的监测、预测仍在循环，而大数据则如同动滑轮，可以帮助减轻负载的重量，能够在突发公共卫生事件前期，提高监测、预测工作的灵活性，面对同样的场景需求可以通过更少的人力、物力、财力达到平衡。基于大数据赋能的重要作用，下列内容进一步阐述大数据在监测、预测各个环节赋能的具体方式。

（一）大数据对突发公共卫生事件预测、监测的赋能方式

“滑轮组”模型（图2）中的“定滑轮”——突发公共卫生事件预测、监测端呈现出“风险监测——信息采集——分析研判——提供预测”的过程逻辑。在面对突发公共卫生事件的场景需求时，随着灾害类型由单一类型灾害发展为复合类型灾害，对预测、监测的专业能力、部门协调、管理工具的要求更高，仅依靠管理层面的队伍建设、组织变革、机制创新、制度优化，难以较好地提高预测、监测能力。“动滑轮”大数据技术端通过“数据获取——数据处理——数据应用——数据赋能评价”的推进，与预测、监测环节结合、双向牵引，能够为有关部门提供坚实的技术基础，使其面对多类型的突发公共卫生事件时做出符合需求的实时监测、精准预测，从而在多次实践中提高预测、监测能力。

（二）大数据与预测、监测的互动

国家大数据局、中央网信办等17部门联合印发的《“数据要素×”三年行动计划（2024—2026年）》中提到，突发公共卫生事件的监测和预测是一个复杂的、多层次的系统工程，但大数据可以提供强大的信息和技术支撑，通过全面整合多渠道的数据实现对疫情的实时监测，从而为决策者提供深度的传播路径分析和感染病例溯源^①。

^① 吴江，陶成煦．激活数据要素赋能千行万业——《“数据要素×”三年行动计划（2024—2026年）》政策解读[J/OL]．情报理论与实践，1-6[2024-02-15]．<http://kns.cnki.net/kcms/detail/11.1762.g3.20240131.1907.002.html>．

1. 多源数据获取和整合

在突发公共卫生事件的监测和预测中，更丰富的信息来源能够为决策者提供更全面、立体的疫情画像，帮助决策者多角度地了解疫情，更有针对性地制定应对措施。通过大数据平台，决策者能够获取多渠道、多主体的信息，如表1所示。数据的多源性也反向驱动大数据平台成为一个多机构、多部门共同参与数据采集和模型建设的协同工作生态系统，这种协同优化机制能够汇聚更丰富的数据和更专业的知识，在突发公共卫生事件的潜伏过程中提供全面性、专业性的监测数据和预测结果。

表1 多源数据类别及内容

Table 1 Categories and contents of multi-source data

数据类别	数据的内容
医疗机构数据	各大医疗机构的实时报告、患者病历和检测数据等关键的疫情传播趋势和病例分布情况
社交媒体数据	社交媒体平台上的言论、评论和实时消息等能够感知社会态度和行为的信息
搜索引擎数据	国内常用的百度、夸克等搜索数据，包含大众关心的健康问题和相关问题的用户端的定位
移动设备数据	溯源感染病例、疫情传播路径的移动设备位置数据
实时新闻报道	各大新闻媒体对疫情的地理分布、最新统计数据、卫生部门的应对措施的报道
公共交通数据	车辆运行情况、客流量等能够评估人员流动情况和传播风险可能

2. 多源数据获取和整合

突发公共卫生事件的预测是建立在对其全面即时的监测基础之上的，而信息的实时反馈是大数据在突发公共卫生事件监测和预测中的重要特征之一。来自各方的信息可以被及时捕捉、高效整合，并通过先进的数据可视化技术将信息以实时疫情地图、感染趋势图、高风险区域的热力图^①形式实时呈现在决策者面前。在展示实时信息的基础上，大数据系统能够设定数据警戒和预警标准，

① 王秉，白明皓. 大数据环境下突发公共卫生事件风险信息感知与分析模型 [J]. 情报杂志, 2021, 40 (11): 176-181.

当数据达到或超过这些标准时，系统能够立即发出预警信号。这种机制有助于在疫情暴发初期即刻启动应急措施，提高对潜在风险的敏感度。

3. 自动化和智能化应用

首先，数据处理自动化。在信息获取过程中，通过自动化的数据采集平台，收集和抓取突发公共卫生事件线上、线下海量数据，为监测和预测提供了原始素材。在处理过程中，通过对数据的去重、转换、清洗、分类、抽取和集成，数据自动提取、数据挖掘、分词技术、语义分析、自动图像识别等大数据技术处理，系统能够从突发公共卫生事件演化过程产生的海量数据中，挖掘出结构化或非结构数据所蕴含的有价值信息^①。

其次，模型训练与优化自动化。利用历史数据、数据特点和需求自动选择合适的机器学习，并让模型能够理解数据的模式、关联、趋势和规则。在预测时，模型保持实时更新，根据实时数据和模型反馈，自动调整模型的参数和结构，从而确保模型能够反映事件的最新状态，提高预测准确性和稳定性。

最后，决策支持与应急响应自动化。利用机器学习算法，建立智能化决策支持系统，帮助决策者快速理解疫情发展态势和可能的影响因素。基于智能分析的结果，自动为决策者生成决策报告和应对方案，展示突发公共卫生事件的关键数据、趋势分析和风险评估，内容包括疫情传播情况、患者分布情况、病毒变异趋势等。同时，根据实时监测数据和预测结果，建立智能化的应急响应流程和机制。在决策和应急响应的过程中，决策者和相关专家只需要分析决策报告和应对方案的可行性，减少人为干预和反应时间，以确保及时性和有效性。

4. 定期评估优化

依托大数据的高度处理、数据挖掘和机器学习的算法优势，持续对突发公共卫生事件实际发展情况与大数据模型预测结果进行对比分析。通过比较预测结果与实际情况的差异，识别出模型的预测偏差和不足之处，进而调整模型参数，如调整模型的学习率、正则化参数、层数和节点数等，以优化模型的拟合能力和泛化能力；改进特征工程，如探索新的特征选择、提取和组合方法，以

^① 周芳检. 大数据时代的重大突发公共卫生事件预警创新[J]. 云南民族大学学报(哲学社会科学版), 2020, 37(5): 114-120.

提高模型对数据的表征能力和预测精度；更新模型选择，对于经过评估后无法满足需求的模型需要进行更换或更新以提高其性能。同时，建立实时监控系统，对模型的运行情况进行持续监测和反馈，及时发现模型性能下降或其他异常情况，并采取相应的措施进行调整和优化，以确保系统能够更迅速、精准地捕捉突发公共卫生事件的发展趋势，以达到最佳预测性能，为决策者提供更为科学、及时的预警和决策支持。

（三）场景需求推动预测、监测能力提高

一是需求导向能够明确数据范围。大数据具有多样性和复杂性，在突发公共卫生事件监测和预测中，实际需求明确了哪些数据类型是最关键和有用的，能够避免陷入数据的过度收集和处理，从而浪费有限的资源和时间。

二是需求导向能够推动技术研发和应用。通过深入了解场景需求，可以明确技术应用的目标、要求、使用方式和技术研发的优先级，进而有针对性地选择适合的技术。同时，在突发公共卫生事件的不同阶段和不同情境下，需求可能会发生变化和调整。以需求为牵引，能够根据实际情况灵活调整监测和预测的策略和方法，保持监测和预测的实效性和实用性。

三是需求导向促进合作和协同。在突发公共卫生事件监测和预测中，各个部门、机构和专业领域可能涉及不同的监测指标、数据来源和分析方法。通过明确需求和准确定义，可以统一各方对于监测和预测的理解和期待，实现信息共享和资源整合。

四、大数据赋能突发公共卫生事件预测、监测的优化

首先，要强化数据采集标准，辨别信息真伪。政府、医疗机构等场所需要对采集、管理、使用个人健康数据的全过程形成统一的标准，保证数据真实度，方便相关组织、机构对数据进行汇总和分析^①。在判定信息真实性过程中，要实

^① 谢文澜，孙雨圻. 基于大数据的公共卫生事件精准应对策略探讨 [J]. 医学与社会, 2021, 34 (6): 119-123.

时关注舆情发展，避免涉疫谣言使民众的认知出现偏差，同时利用情感特征、传播特征、时序特征、文本预测特征等方式进行虚假信息识别。在确定信息真实性后，及时利用官方权威信息平台进行辟谣，以免涉疫谣言传播范围进一步扩大引发社会矛盾和公众负面情绪。

其次，要完善数据资源体系，优化数据流通环境。加快开展行业共性数据资源库建设，建立数据共享机制，推动政府数据开放平台、企业、公众形成循环闭环，简化数据整合的流程，降低数据整合的难度和成本。建立开放式的数据平台和数据市场，为数据提供者 and 使用者提供便捷的数据交易和共享平台，发挥数据要素乘数效应，促进数据创新和应用。

再次，要保护个人健康数据安全。对于不需要识别个人身份的敏感医疗健康信息，需要采用匿名化处理，以保护个人隐私^①。在数据储存期间，需要采用加密存储和备份技术，防止数据丢失或遭受恶意攻击。还需定期对存储系统进行安全审计和漏洞修复，以及时发现和排除安全隐患。同时，要建立严格的访问权限控制机制，限制数据的访问范围和权限，对数据使用记录进行实名监测和记录。根据用户的身份和需求对不同级别的数据访问权限进行划分，将数据泄露和滥用的风险降到最低，确保数据的合法使用和安全保护。

最后，要建立长效技术创新机制^②。保持持续的研发投入，制定明确的技术迭代和升级计划，避免技术停滞，并鼓励相关机构和个人持续推动大数据技术的优化和创新。

五、结论

突发公共卫生事件的不可预测性和复杂性对现有的监测和预测体系提出了严峻挑战。在大数据技术的赋能下，构建智能化、自动化的监测和预测系统，不仅能显著提高效率和精准性，还能为决策提供更科学的依据。本文基于“滑

① 祝丙华，王立贵，孙岩松，等. 基于大数据传染病监测预警研究进展 [J]. 中国公共卫生, 2016, 32 (9): 1276-1279.

② 陈茁，陈云松. 应急管理研究中的定量方法：数据、因果和算法 [J]. 中国行政管理, 2023, 39 (11): 139-147.

轮组”模型深入探讨了大数据在突发公共卫生事件监测和预测中的应用与优化策略。研究表明，大数据通过多源数据的实时获取、精准处理、自动化分析和智能化决策支持，能够在潜伏期和征兆期显著提升监测和预测能力，从而为公共卫生风险治理和应急管理提供坚实的技术支撑。

然而，大数据赋能的监测与预测仍面临一系列挑战，如虚假信息的泛滥、数据孤岛现象的存在、个人隐私保护的不足，以及技术创新的可持续性问题。因此，本文提出以下优化方向：首先，强化数据采集标准，利用技术手段辨别信息真伪，降低虚假信息对监测结果的干扰；其次，完善数据资源体系，优化数据流通环境，建立跨部门、跨机构的数据共享机制，推动形成循环闭环的生态；再次，加强对个人健康数据的保护，采用匿名化处理、加密存储及分级访问权限等多重措施，确保数据安全；最后，建立长效技术创新机制，持续推动大数据技术的迭代升级，确保其在应急管理中的前沿性和实效性。

通过需求导向驱动大数据技术的场景化应用，可以更精准地定义数据类型和技术优先级，并实现高效的跨部门协作和资源整合。未来，大数据技术应进一步与人工智能、云计算等前沿技术融合，构建更加完善的监测和预测体系，为应对突发公共卫生事件提供智能化的全流程支持，为全球公共卫生安全治理提供中国方案和智慧借鉴。

Algorithm Optimization of Big Data in Predicting and Monitoring Public Health Emergencies

Weng Xinrui

Beijing University of Chemical Technology, Beijing

Abstract: The complexity and unpredictability of public health emergencies pose significant challenges to traditional monitoring systems. This article proposes the “pulley system” model, leveraging big data technologies to achieve risk monitoring, information collection, analysis and evaluation, and predictive support, thereby enhancing decision-making efficiency and accuracy. To address issues in current data applications, such as difficulties in identifying misinformation, data silos, privacy breaches, and insufficient sustainability of technological innovation, optimization directions are proposed. These include strengthening data collection standards, improving data resource systems, ensuring the security of personal health data, and establishing long-term mechanisms for technological innovation. Furthermore, the article highlights the application of big data driven by scenario-specific needs, enabling real-time integration and analysis of multi-source data to advance the intelligence and automation of epidemic prediction and monitoring. By summarizing practical experiences and theoretical models, this article provides a systematic approach to empowering public health emergency monitoring with big data, offering insights for innovation in risk governance and emergency management.

Key words: Big data; Epidemic monitoring; Risk prediction; Pulley system