

重大公共卫生事件环境下多角色画像系统模型研究^{*}

张红伟 郭莹 陈玲 任淑敏

(济宁医学院图书馆 济宁 272067)

〔摘要〕 介绍我国疫情防控研究概况及相关理论、方法,从架构、功能等方面详细阐述多角色画像系统模型构建,指出该模型能够为疫情发布、监测、排查、分析、预警、防控等提供数据全景支撑,为传染病疫情防控提供新思路。

〔关键词〕 用户画像;新冠肺炎;大数据技术;手机定位

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2021.03.003

Study on the Multi-role Portrait System Model in the Context of Major Public Health Events ZHANG Hongwei, GUO Ying, CHEN Ling, REN Shumin, Library of Jining Medical University, Jining 272067, China

〔Abstract〕 The paper introduces the study situation of epidemic prevention and control in China and relevant theories and methods, expounds the building of multi-role portrait system model in detail from the aspects of architecture, functions and others, points out that this model can provide data panoramic support for the epidemic report, surveillance, screening, analysis, early warning, prevention and control, etc., and provide new ideas for the epidemic prevention and control of infectious diseases.

〔Keywords〕 user portrait; COVID-19; big data technology; mobile phone positioning

〔修回日期〕 2020-08-25

〔作者简介〕 张红伟,硕士,副研究馆员,发表论文 10 余篇;通讯作者:任淑敏,教授,发表论文 50 余篇。

〔基金项目〕 济宁医学院国家社会科学基金培育项目“数字人文视域下高校图书馆服务功能拓展研究”(项目编号:JYP2018SK08);济宁医学院医学人文教育研究专项课题“创新视域下图书馆对医学生人文素质培养的路径探析”(项目编号:201822);济宁医学院教师科研扶持基金“基于数字人文的医学高校图书馆功能与服务路径探究”(项目编号:JYFC2018RW007)。

1 引言

2019 年 12 月 8 日湖北省武汉市发现第 1 例新型冠状病毒肺炎患者,之后迅速在全国范围内蔓延。习近平总书记做出重要指示:把人民群众生命安全和身体健康放在第 1 位,努力做好疫情监测、排查、预警等工作,及时准确、公开透明发布疫情^[1]。新冠肺炎疫情防控工作在全国范围内全面展开。我国关于疫情监测^[2-3]、分析^[4-5]、预警^[6-7]等研究特点包括以下几方面:一是多针对流行性感冒、手足口病、艾滋病等传播范围较小的病种;二是研究周期跨度较长,多是在监测多轮疫情前提下,疫情防控工作结束后进行的分析与预警。在波

及全国甚至世界范围的重大传染病疫情暴发期间,疫情监测、发布、分析与预警研究不足。在新型冠状病毒快速传播的情况下如何对疫情相关数据进行有效监测、公布与分析,在分析疫情数据基础上及时预警并制定适当防控措施已经成为亟须解决的问题,对此在真实数据基础上,本文提出勾勒虚拟信息模型的用户画像系统^[8]。

2 相关理论和方法

2.1 大数据技术

目前大数据尚缺乏统一定义,通常被理解为一个数据量庞大的集合。大数据的战略意义不在于海量数据本身,而是通过数据挖掘与分析,厘清其内部蕴藏的逻辑关系和发展规律。大数据技术是采集、存储、分析和处理海量数据,进而发现事物发展规律,预测和防范事件发生的技术总称,主要包括大数据采集、存储、分析和挖掘、可视化技术。

2.2 用户画像

一种刻画用户信息模型的技术,在收集、统计、挖掘、分析大量属性信息和行为数据基础上抽象出用户标签体系,进而勾勒虚拟用户信息模型,凸显用户信息全貌,构建标签体系是用户画像的核心任务。用户画像技术最初应用于市场营销、社交媒体及数字图书馆服务等领域,随着大量研究的持续开展其应用范围不断拓展。张杨燚等通过构建城市画像刻画多维度城市特点,为建设智慧城市提供科学依据^[9];韩梅花等通过提取抑郁特征为抑郁症患者画像,根据抑郁情感指数向患者推送有助治疗的文献资源^[10];杨晓宇等构建患者画像析出患者需求,为临床科室实施精准医疗提供科学决策依据^[11]。借助大数据技术,用户画像利用用户标签体系提供了一种全面呈现事物信息特点、预测事物发展规律并为制定未来发展方案提供数据支撑的简易方法。

2.3 基于手机定位数据的出行轨迹研究

手机定位数据主要包括手机信令、传感器和社交网络数据。随着智能手机的普及,定位数据获取难度降低,个体出行行为特征成为当前研究热点^[12],其中出行轨迹提取是最基本内容。在此基础上手机定位数据研究主要集中于停驻点识别、出行交通量(Origin Destination, OD)信息获取以及出行方式识别等方面。

2.4 地理信息系统(Geographic Information System, GIS)

GIS是一门与计算机、地理信息、遥感及信息科学相关的交叉学科。主要包含空间和属性两种数据,通过对空间数据进行采集、存储、分析处理,集成整合空间数据与属性数据,实现空间实体的物理定位、属性查询、分类汇总与可视化显示等功能。GIS广泛应用于资源调查与管理、灾害预测和预防、环境监测与保护等方面。流行病防控研究领域已引入GIS,程文炜等利用GIS分析中老年糖尿病患者的空间分布规律和空间影响因素,为完善区域化糖尿病防控措施提供科学依据^[13]。查文婷等利用GIS分析手足口病在长沙市不同区域的时空分布特征,为卫生资源优化配置、防控手足口病指明方向^[14]。依据强大的空间计算和分析能力,GIS已成为流行病研究的有效工具,在动态展示流行病分布,依据时空、环境等因素分析与预测流行病发展趋势等方面展现出良好性能。

3 构建多角色画像系统模型

3.1 概述

本系统可以通过相关数据直观展示居民、社区、医院、工作机构以及各级行政区域有关疫情发展的基本状态,为疫情防控提供预警和决策依据。该模型分为3层架构,按照大数据处理流程自底向上分别是基础数据层、数据分析层和信息服层,见图1。

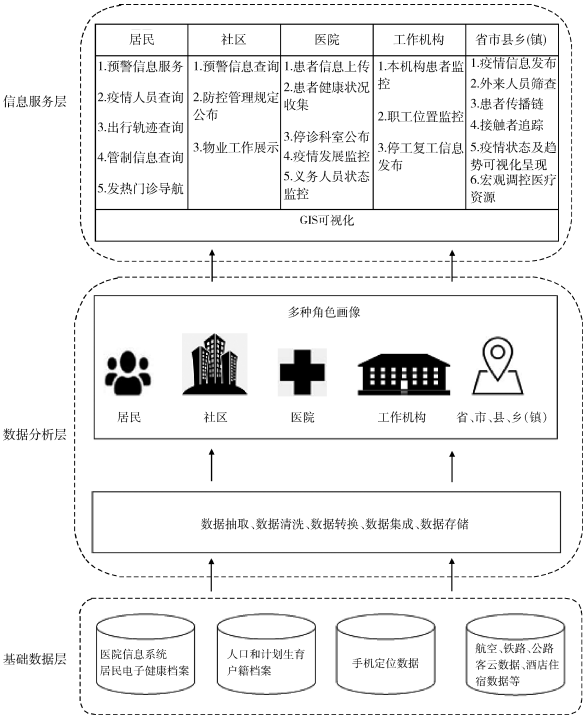


图 1 多角色画像系统模型架构

3.2 基础数据层

系统的大数据来源，涉及疫情防控所需的所有底层数据，主要包括医院信息、居民电子健康档案、人口和计划生育、户籍档案、手机定位、酒店住宿以及航空、铁路、公路客运等。这些数据可反映当前疫情发展基本态势，是疫情分析的基础。

3.3 数据分析层

3.3.1 概述 利用大数据技术抽取基础层数据，经过数据清洗、转换、集成、存储等环节，以标签属性方式输出至信息服务层。标签体系是画像系统模型的重要环节，标签设置以展示疫情环境下各种角色基本状态，为疫情防控提供依据。模型分别创建了居民、社区、医院、工作机构和各级别行政区域标签体系，从不同角度展现当前社会状态，分别从微观和宏观层面反映疫情发展情况。

3.3.2 居民标签体系 居民是新冠肺炎疫情防控工作的基本个体，居民属性主要包括基本信息、健康状况、亲戚关系、籍贯与常住地、出行信息 5 个维度。基本信息主要包括姓名、性别等静态信息；

健康状况不仅能体现居民是否为新冠肺炎患者，还能提取其既往病史、门诊病例等，为制定治疗方案提供依据；亲戚关系、家庭住址和工作单位信息可以体现居民主要接触人员范围，为向相关人员提供预警信息和寻找密切接触者提供线索；出行信息体现居民近期活动范围，对于确诊患者而言可以结合患者回忆找出感染后去过的地方和接触人员，向密切接触者和相关区域负责人发送预警信息；对于普通居民而言可以显示是否与确诊患者和疑似患者出行轨迹有交叉，为做好预防措施提供依据，见表 1。

表 1 居民标签体系

1 级标签	2 级标签
基本信息	姓名
	性别
健康状况	年龄
	工作单位
	身份证号码
	手机号码
	是否新冠肺炎确诊患者
	是否新冠肺炎疑似患者
	是否新冠肺炎密切接触者
	既往病史
	门诊病例

续表 1	
亲戚关系	电子健康档案
	家庭成员
	主要亲戚成员
籍贯与家庭住址	籍贯
	常住地
	具体家庭住址（精确到社区或村等）
近期出行信息	航空、铁路、公路客运乘坐信息
	酒店住宿信息
	所有记录身份证号码的公共场所信息
	近期出行轨迹

3.3.3 社区标签体系 社区是新冠肺炎疫情区域防控中的重要环节，是居家隔离政策的执行主体。社区标签体系，见表 2。

表 2 社区标签体系	
1 级标签	2 级标签
社区基本信息	居民总数
	地理位置
	楼宇数量
	本社区防控管理规定
新冠肺炎确诊患者信息	数量
	基本信息
	所在楼宇、单元及门牌号
	所在医院
新冠肺炎疑似患者信息	数量
	基本信息
	所在楼宇、单元及门牌号
	所在医院
尚在医学观察的密切接触者信息	数量
	基本信息
	所在楼宇、单元及门牌号
	隔离情况
解除医学观察的密切接触者信息	数量
	基本信息
	所在楼宇、单元及门牌号
	隔离情况

3.3.4 医院标签体系 医院是检测和救治新冠肺炎患者的主体，是疫情防控的主战场。整个医院标签体系综合体现了医院新型冠状病毒的接诊压力，为疫情防控提供依据，见表 3。

表 3 医院标签体系	
1 级标签	2 级标签
医院基本信息	相关医护人员数量
	床位数量
	疫情期间正常运行的科室
	疫情期间停诊科室

续表 3	
医疗资源信息	医疗设备信息
	医疗防护装备信息
	所需药品信息
	收治新冠肺炎
患者信息	在院患者总数
	在院患者信息
	重症患者总数
	重症患者信息
发热门诊就诊信息	治愈患者总数
	治愈患者信息
	死亡患者总数
	死亡患者信息
	就诊人数总量
	每日就诊人数总量
	各医生接诊人数总量
	各医生每日接诊人数总量

3.3.5 工作机构标签体系 工作机构在疫情期间仍然承担生产任务，是疫情防控的基本单元。工作机构属性与社区属性相似，主要包括基本情况、确诊患者信息、疑似患者信息、尚在和解除医学观察的密切接触者信息 5 个维度。基本情况包含本机构人员总数及地理位置等信息；确诊患者、疑似患者及密切接触者体现了本机构人员有关疫情的基本情况，为疫情防控提供基本依据，为停工和复工提供数据参考。

3.3.6 各级行政区域标签体系 各级行政区域是各级行政机构宏观防控政策的执行客体，其属性与社区属性相似，主要包括辖区基本信息、确诊患者信息、疑似患者信息、治愈患者信息、死亡患者信息、尚在和解除医学观察密切接触者信息及管制信息 7 个维度。辖区基本信息包括人口总数、土地面积、定点医院等；管制信息显示本区域基本的道路、区域管制情况。

3.4 信息服务层

3.4.1 功能 及时公布画像系统模型中各种角色疫情信息，结合 GIS 以可视化形式展示，依据用户权限为居民、社区帮扶与物业工作人员、医院医护人员、工作机构管理人员及职工、各级管理人员与疫情防控人员等提供有关新冠肺炎疫情相关信息的

查询与获取服务,对与感染人员和密切接触者有联系的人员进行及时预警,为下一步防控工作安排提供数据支撑。信息服务层从微观和宏观不同层面,基于居民、社区、医院、工作机构、各级别行政区域等不同角度,以 GIS 可视化方式通过大数据展示社会各界疫情发展状态和趋势,为疫情防控精准施策提供科学依据。

3.4.2 居民层面 一是预警信息服务。一旦发现新冠肺炎确诊患者、疑似患者或密切接触者,系统马上向其家庭成员、主要亲属、工作单位负责人、社区同一楼宇居民或同村居民发送预警信息,提醒相关人员避免与其接触,做好防护措施等;向与其近期密切接触人员发送预警信息,提示需要做医学检测或居家隔离等。二是疫情相关人员信息查询。在地图中呈现用户居住地周边确诊患者、疑似患者和密切接触者等各类人员信息,显示患者所在社区及距离等位置信息,提醒居民避开感染严重区域,降低被感染几率。三是出行轨迹查询。用户可以查询近期周边确诊患者或疑似患者出行轨迹,明确是否与自己的出行轨迹有交叉,根据具体情况做好预防措施。四是管制信息查询。查询周边超市、药店等是否正常营业,某道路或区域是否交通管制等,为出行做好准备。五是发热门诊导航。为用户提供周边发热门诊信息。

3.4.3 社区层面 患者信息查询,为社区帮扶人员、物业管理提供及时、动态的本社区及周边社区新冠肺炎确诊患者、疑似患者和密切接触者数量及住址信息,为社区制定防控措施提供参考。物业管理人员及时公布社区防控管理规定,帮助居民提前做好相应准备。动态展示物业清洁和消毒工作,增强社区居民防疫信心。

3.4.4 医院层面 一是及时上传新冠肺炎患者信息。患者一旦确诊或被确定为疑似病例,立即将信息上传到系统中,以便追踪与该患者相关的密切接触者。二是根据新冠肺炎患者身体健康状况制定诊疗方案。患者画像描述其既往病史、门诊病历及电子健康档案等,根据这些诊疗信息结合患者具体情况制定治疗方案,为全力抢救患者提供依据。三是及时公布疫情期间停诊科室,开通网上门诊,避免

大量人员进入医院,切断传播途径。四是综合审视医院收治患者数量、周边疑似患者数量、医疗资源,宏观评估医院医疗压力,必要时争取相关支援。五是监控医护人员工作状态。疫情期间,医护人员工作劳动强度剧增,严格监控各医护人员工作时长及身体状况,确保医护人员身体健康。

3.4.5 工作机构层面 一是时刻关注系统中本机构职工有无新冠肺炎确诊患者、疑似患者或密切接触者,一旦出现相关人员,针对具体情况立即制定防控方案,防止疫情在本单位传播。二是密切关注本单位所有人员具体位置,根据疫情分布情况,一旦发现身处疫情严重地区职工立即执行隔离方案。三是及时公布停工和复工部门。根据疫情发展明确各部门停工和复工时间并及时公布,提前让相关职工做好准备。

3.4.6 各省、市、县、乡(镇)等行政区域层面 一是及时公布疫情信息,包括确诊患者、疑似患者、密切接触者等人员信息以及交通路线及区域管制情况,扩大宣传力度,做好疫情防控安排。二是时刻监控进入辖区外来人员,严格筛查新冠病毒携带者。根据手机定位数据查看进入辖区人员近期是否去过疫情严重地区,一旦发现相关人员则采取禁止进入辖区或隔离等措施。三是绘制确诊患者传播链,一旦发现确诊患者将其加入患者传播链,分析各患者之间关系,为寻找潜在被感染患者提供线索。四是追踪密切接触者,做好隔离工作。一旦发现确诊患者或疑似患者,利用系统追寻密切接触者并采取有效隔离和预防措施。五是可视化展示疫情分布、各类患者居住地分布及增长趋势,制定有效管制策略。六是可视化展示各医院收治新冠肺炎患者数量分布,宏观调配医疗资源。

4 结语

新冠肺炎疫情防控已成为我国当前重点工作。本文构建多角色画像系统模型,用以全方位反映居民、社区、医院、工作机构、各级行政区域等有关社会各界的疫情发展状态,结合 GIS 以可视化方式呈现给用户,为疫情公布、监测、排查、分析、预

警、防控等提供数据全景支撑。该系统为新冠肺炎及类似传染病疫情防控提供了新思路。

参考文献

- 1 中国新闻网. 习近平主持中央政治局常委会会议 研究疫情防控工作 党中央成立应对疫情工作领导小组 [EB/OL]. [2020 - 01 - 25]. <https://www.chinanews.com/gn/2020/01-25/9069630.shtml>.
- 2 焦玲艳, 王颖, 韩啸宇. 2013 - 2018 年天津市滨海新区流感监测分析 [J]. 现代预防医学, 2019, 46 (19): 3612 - 3615.
- 3 李多, 徐闻, 伏晓庆, 等. 2017 年云南省乙型流感病毒监测结果及其血凝素基因特征分析 [J]. 病毒学报, 2019, 35 (4): 606 - 611.
- 4 范家诚, 姚永明, 郑敏, 等. 2011 - 2016 年贵州省少数民族艾滋病疫情流行特征分析 [J]. 中国艾滋病性病, 2020, 26 (1): 40 - 42.
- 5 张桂欣, 许玉成, 周志峰, 等. 2008 - 2018 年深圳市福田区流行性感冒聚集性疫情流行特征分析 [J]. 现代预防医学, 2019, 46 (19): 3471 - 3474, 3486.
- 6 祝姗姗, 高洁. 基于动态网络生物标志物的甲流疫情早期预警算法研究 [J]. 病毒学报, 2018, 34 (1): 9 - 15.

- 7 王思嘉, 聂绍发. 湖北省潜江市手足口病疫情不同预警模型效果比较 [J]. 中国公共卫生, 2017, 33 (12): 1760 - 1763.
- 8 魏明珠, 张海涛, 刘雅姝, 等. 多维属性融合的社交媒体高影响力人物画像研究 [J]. 图书情报知识, 2019 (5): 73 - 79, 100.
- 9 张杨斌, 谢辉, 毛进, 等. 面向城市数据画像构建的多源数据需求与融合方法研究 [J]. 情报理论与实践, 2020, 43 (6): 88 - 96.
- 10 韩梅花, 赵景秀. 基于“用户画像”的阅读疗法模式研究——以抑郁症为例 [J]. 大学图书馆学报, 2017, 35 (6): 105 - 110.
- 11 杨晓宇, 王秀民, 李黎. 以患者画像识别技术为核心的 CRM 系统实践精准医疗 [C]. 珠海: 2016 中华医院信息网络大会, 2016: 1 - 4.
- 12 陈旭, 郑浩毅. 基于手机定位数据的个体出行行为特征分析综述 [J]. 综合运输, 2019, 41 (9): 38 - 44.
- 13 程文炜, 闫晓芳, 史静琚, 等. 基于 GIS 的中国中老年人糖尿病空间分布及其影响因素分析 [J]. 中华疾病控制杂志, 2017, 21 (11): 1082 - 1087.
- 14 查文婷, 王玲, 郑剑, 等. 基于 GIS 技术分析长沙市 2006 - 2013 年手足口病流行病学特征分析 [J]. 中华疾病控制杂志, 2017, 21 (4): 345 - 348.

(上接第 13 页)

4 结语

领域知识图谱是支撑智能应用的重要基础。本研究利用自然语言处理技术及计算机辅助人工构建的方式, 从新冠肺炎相关科技文献、诊疗方案等多来源发现语义知识, 从等级体系、主题分类、关系类型等多维度进行深度组织, 探索出一种专题知识图谱构建方法, 初步建成的新冠肺炎知识图谱提供新冠肺炎相关概念等级分类、主题分类及病因、传播、诊疗、防护等多种语义关联, 可用于支撑生物医学文献主题标引、语义标引及主题分类, 以及文献信息服务系统的主题、语义、分类检索等应用。今后的工作重点包括新知识发现及补充完善、与 OpenKG 中新冠肺炎开放图谱的知识融合、基于所建图谱的知识推理等方面。

参考文献

- 1 奥德玛, 杨云飞, 穗志方, 等. 中文医学知识图谱 CMeKG 构建初探 [J]. 中文信息学报, 2019, 33 (10): 1 - 9.
- 2 中国医学科学院医学信息研究所. 新型冠状病毒肺炎术语集 [EB/OL]. [2020 - 11 - 20]. <https://www.ncmi.cn/phda/dataDetails.do?id=CSTR:A0006.11.A0003.202002.000590-V1.0>.
- 3 Schriml L M, Arze C, Nadendla S, et al. Disease Ontology: a backbone for disease semantic integration [J]. Nucleic Acids Res, 2012 (40): D940 - D946.
- 4 孙海霞, 李军莲, 李丹亚, 等. 基于 CMeSH 语义系统的领域自由词 - 主题词语义映射研究 [J]. 现代图书情报技术, 2013 (11): 46 - 51.
- 5 Pavel Shvaiko, Jerome Euzenat. A Survey of Schema - based Matching Approaches [J]. Journal on Data Semantics IV, 2005 (4): 146 - 171.