

防灾减灾知识服务系统及其应用研究

王卷乐^{1,2*}, 韩雪华^{1,2}, 卜坤³, 张敏^{1,2}, 王晓洁^{4,1}, 袁月蕾¹

1. 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室, 北京 100101;
2. 中国科学院大学资源与环境学院, 北京 100049;
3. 中国科学院东北地理与农业生态研究所, 长春 130102;
4. 山东理工大学建筑工程学院, 淄博 255049

摘要: 防灾减灾知识服务系统是在联合国教科文组织 (UNESCO) 防灾减灾使命驱动下, 由依托中国工程院的 UNESCO 二类中心国际工程科技知识中心支持建立的。该系统以灾害元数据标准为切入点, 汇聚和生产防灾减灾数据产品, 利用大数据技术和地理信息技术实现防灾减灾知识应用, 并向全球用户开放使用。防灾减灾知识服务系统已建立了 16 项在线知识应用, 形成了基于遥感大数据、网络大数据进行防灾减灾信息挖掘、分析和可视化的应用能力。2019 年末在中国武汉突发新型冠状病毒肺炎疫情, 当前中国乃至全球都面临着新型冠状病毒肺炎疫情防控的严重挑战。针对这一形势, 防灾减灾知识服务系统快速构建灾害舆情分析在线知识应用。该知识应用基于中国新浪微博大数据, 从数以百万计的相关文本记录中获取其空间位置和相关话题语义信息。基于核密度分析等方法, 形成疫情演变信息的可视化分布图, 并通过在线知识应用发布共享。基于隐含狄利克雷分布和机器学习算法, 分析疫情相关的微博时空分布和公众情绪, 为新型冠状病毒肺炎疫情防控提供信息和应用支撑。

关键词: 防灾减灾; 知识服务; 社交媒体; 舆情分析; 新型冠状病毒

DOI: 10.3974/geodp.2020.01.03

1 前言

灾害的减除和预防是全球共同面临的紧迫课题。联合国教科文组织 (UNESCO) 长期重视这一领域的全球合作, 并在其自然科学部中设置有地球科学与地质灾害风险减除部门 (Earth Sciences and Geo-hazards Risk Reduction)。国际工程科技知识中心 (IKCEST) 是 UNESCO 于 2013 年依托中国工程院建立的一个二类中心。结合其定位和使命, UNESCO 于 2015 年寻求与 IKCEST 在防灾减灾方面的密切合作。在 UNESCO 防灾减灾的使命驱动下, IKCEST 于 2016 年启动了防灾减灾知识服务分中心建设, 委托中国科学院地理科学与资源研究所承建该分中心, 并开发上线运营防灾减灾知识服务系统。

收稿日期: 2020-02-15; 修订日期: 2020-02-28; 出版日期: 2020-03-25

基金项目: 中国工程科技知识中心建设项目 (CKCEST-2020-2-4); 中国-巴基斯坦地球科学研究中心 (Y99R0900F2); 中国科学院 (XXH13505-07)

*通讯作者: 王卷乐 R-8881-2016, 中国科学院地理科学与资源研究所, wangjl@igsnr.ac.cn

引用格式: 王卷乐, 韩雪华, 卜坤等. 防灾减灾知识服务系统及其应用研究[J]. 全球变化数据学报, 2020, 4(1): 16–23.
DOI: 10.3974/geodp.2020.01.03.

2 防灾减灾知识服务系统

2.1 目标与愿景

防灾减灾知识服务系统的目标是面向全球防灾减灾需求,在 UNESCO 防灾减灾的使命驱动下,制定灾害元数据国际/国家标准或最佳实践,在统一标准体系下构建全球灾害元数据库;以中国及其周边地区和世界典型地区的主要灾种为对象,整合和集成国家/区域尺度的灾害数据和信息,建立灾害数据库;以大数据挖掘和分析技术为支撑,建立防灾减灾知识服务系统平台,挖掘灾害的预防、救援、重建、评估等方面的数据库开发方法和知识信息服务模式;广泛开展防灾减灾专题服务、教育传播、国际培训和合作交流,发挥对 UNESCO 防灾减灾工作的支撑作用^[1]。

防灾减灾知识服务系统的愿景是为当前全球减灾防灾提供平台、技术、数据、教育、知识等方面的知识服务,积累防灾减灾数据库、产品库、知识库等科技和学术资源,打通防灾减灾国内与国际资源联通,凝聚国际防灾减灾典型案例和应用示范,支撑“一带一路”倡议下的区域防灾减灾应用,成为 UNESCO 防灾减灾国际合作的重要基础和支点,显著提升 IKCEST 的国际影响力。

2.2 系统架构

防灾减灾知识服务系统在技术架构上优先采用开放的国际技术标准和开源的 Web 技术,采用边应用边开发的迭代开发模式,实现模块化机制按需扩展信息服务平台,使得用户可快速获取数据、地图、文献、视频等各类灾害知识资源和专题知识服务。系统整体架构如图 1 所示。

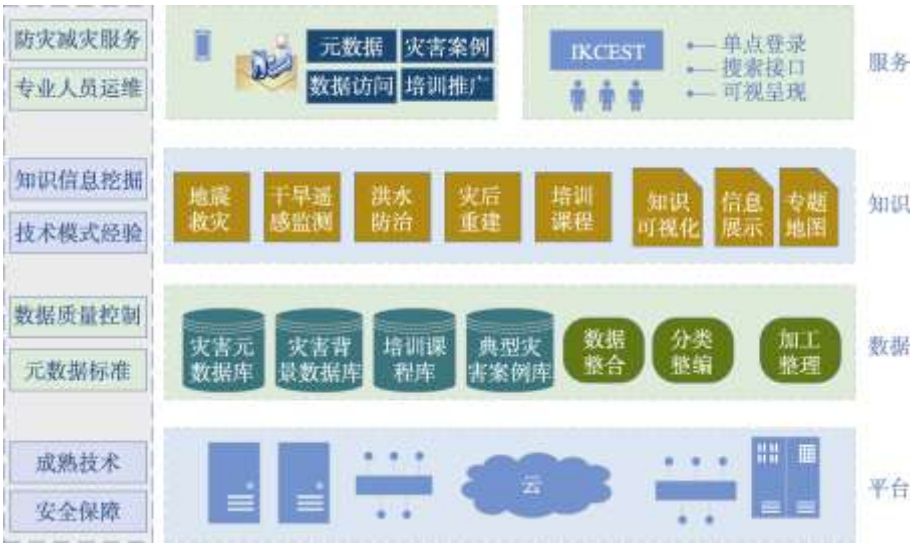


图 1 防灾减灾知识服务平台架构

底层的数据资源存储方案采用阿里云模式,构建文件服务器、元数据库服务器、数据库服务器、地图服务器以及用于解析前端用户访问的 Web 服务器。在一系列开放 Web 技术的支持下,实现数据录入、信息发布、权限管理等编辑和运维功能,以及地图可视化、文

献全文检索、用户行为分析和多灾害专题标签过滤等功能,支撑针对灾害机构分布、灾害地图浏览和防灾减灾专题应用的知识应用功能。

防灾减灾知识服务系统开发基于 B/S 模式,采用 Python + Tornado + TorCMS 应用框架。系统前端采用 HTML5 + CSS3 技术,并使用 JQuery、Bootstrap 3 框架,后台开发语言为 Python 3.4 以上版本。数据持久化使用 PostgreSQL 数据库,核心属性映射于数据库字段,扩展属性则存储与 PostgreSQL 的 JSONB 字段中。用于灾害地图和空间数据可视化的 WebGIS 后台使用 MapServer 地图服务器,前端分别使用 Leaflet 与 OpenLayers 3 JavaScript 库。元数据管理采用 pycsw 开发。pycsw 是 OGC 网络编目服务标准的 Python 语言实现,该服务标准定义了一套统一的接口,用于对空间信息及相关数据进行检索、查询和浏览^[2]。

2.3 产品体系

(1) 元数据标准和技术规范。以元数据标准为基础,在灾害核心元数据标准的研究基础上扩展灾害专家库、机构库、视频课件等多类资源的元数据标准。制定防灾减灾数据管理、防灾减灾开放服务等灾害数据管理与服务技术规范。

(2) 全球灾害元数据库。基于互联网中全球与灾害相关的平台和专业数据库,采用网络爬虫技术获取全球的地震、干旱、洪涝、台风、森林火灾、高温热浪等灾害元数据信息,采用自然语言处理、信息抽取等技术,完成灾害元数据信息的分词、过滤、关键词提取等处理,并结合受控词表,完成语义标签提取和灾害元数据分类。

(3) 中国灾害地图图件库。收集自然灾害地图,进行扫描、整理,并基于地理信息专业技术平台进行处理,形成灾害专题地图图件数据库。对加工处理完成后的地图图件资源需进行编码,保证每一幅地图图件编号的唯一性。

(4) 专题灾害数据库。基于遥感对地观测、历史统计资料挖掘等多种手段自主构建灾害专题数据集。例如,基于面向对象的遥感解译方法获得中蒙铁路沿线的土地退化数据产品;基于历史气象站点监测数据,生产区域月度历史气象灾害数据产品;获取和集成自 1949 年以来形成包括北京、上海、重庆等地在内的典型特大城市地区灾害数据集等。

(5) “一带一路”灾害数据库。集成“一带一路”沿线 65 个国家和地区孕灾背景数据,主要包括基本国情、自然资源、政治经济三大类。基于遥感数据和网络资源,搜集、挖掘和整编中巴经济走廊高温热浪、洪水、地震的损失程度数据。基于遥感和气象观测等数据,获取中俄毗邻地区暴雨洪涝风险分布数据集、中俄毗邻地区极端降水事件数据集等。

(6) 面向 SDGs 的支撑专题数据库。针对 SDG15 中的森林管理、土地退化等重大资源环境和生态灾害问题,采用遥感和地球大数据的技术手段,获取针对 SDG15.1 和 SDG15.3 的数据集产品。例如,中国区域森林分类数据产品,蒙古国荒漠化数据产品,黄河三角洲土地盐渍化退化数据等。

(7) 通用资源建设。参照相关数据规范标准,持续开展灾害专家库、灾害机构库、灾害文献库、灾害课件视频库、全球工程案例等内容建设,并通过信息共享和知识应用工具两种方式向用户开放。

2.4 用户服务

防灾减灾知识服务系统团队利用本平台持续开展用户服务。主要用户对象分为五类：一是 UNESCO 等国际组织或机构，二是防灾减灾相关政府机构和管理技术人员，三是从事防灾减灾的科技工作者，四是高等教育机构师生，五是社会公众。用户访问量达到 1.3 万·月⁻¹，其中有近 50%的用户来自于国际，主要来自于美国、日本、印度、菲律宾、英国等国家。

3 在线知识应用

在线知识应用是防灾减灾知识服务系统提供的一种典型应用模式。它借用于底层的各类数据、地图、机构、专家、视频、课件等资源，以特定应用需求牵引，通过数据集成处理和可视化技术支撑，提供平台前端用户交互和展示。目前已提供 16 个在线知识应用，这些应用可以在防灾减灾知识服务系统首页（<http://drr.ikcest.org>）中找到并访问（图 2）。



图 2 防灾减灾知识服务系统页面展示

防灾减灾知识服务系统目前提供的在线知识应用列于表 1。

表 1 防灾减灾知识服务系统在线知识应用列表

序号	知识应用名称	在线地址	服务功能
1	防灾减灾组织机构知识地图服务	http://drr.ikcest.org/app/s8349	获取全球灾害机构,并提供在线可视化和一站式导航服务
2	全球地震分布可视化地图服务	http://drr.ikcest.org/app/s9834	实时通过 USGS 接口,获取全球地震分布数据,并在线可视化展示
3	中国历史灾害地图可视化服务	http://drr.ikcest.org/app/s7834	获取历史地图图件,经扫描纠正等处理后可视化发布,提供编辑功能
4	防灾减灾典型案例的中国和国际经验	http://drr.ikcest.org/case/index.html	收集全球典型案例,从灾前预防、灾中救援和灾后重建等方面展示
5	中国南方森林冰冻雨雪防灾减灾知识应用	http://drr.ikcest.org/knowledge_service/forest.html	利用 Anusplin 软件进行空间离散化处理,并提供可视化服务
6	中国松辽流域洪水灾害防洪抢险知识应用	http://drr.ikcest.org/knowledge_service/control_flood.html	基于 WebGIS 功能,提供洪水灾害数据和信息空间分布展示和分析服务
7	“一带一路”耕地干旱水平时空展示专题知识应用	http://drr.ikcest.org/knowledge_service/drought.html	建立降水距平百分率干旱模型,提供耕地分布的展示和时空序列分析
8	鄱阳湖悬浮物浓度反演逐季空间分布数据服务知识应用	http://drr.ikcest.org/knowledge_service/poyang_lake.html	对鄱阳湖 4 个季节进行数据建模反演,形成多年时空序列可视化分析
9	蒙古高原干旱监测逐年空间分布数据服务知识应用	http://drr.ikcest.org/knowledge_service/mongolian.html	基于 Ts-NDVI 通用特征空间构建稳定的干旱监测模型,实现多年时空序列分析
10	鄱阳湖叶绿素浓度反演逐季空间分布数据服务知识应用	http://drr.ikcest.org/knowledge_service/poyang_yls.html	采用半经验、经验方法获得鄱阳湖叶绿素 a 浓度估算模型,实现可视化分析
11	蒙古国孕灾环境土地覆盖全要素数据服务知识应用	http://drr.ikcest.org/knowledge_service/mongolian_lc.html	利用面向对象的解译方法所得各类型土地覆盖要素的分布,并可视化展示分析
12	中蒙俄经济走廊主要历史灾害分布时空数据服务知识应用	http://drr.ikcest.org/knowledge_service/zmezl.html	收集多源灾害数据和信息,并提供可视化展示和分析
13	寿光洪水灾害公众情绪时空分布知识应用	http://drr.ikcest.org/knowledge_service/shouguang.html	利用微博文本大数据, LDA 主题模型和随机森林算法进行话题抽取与分类
14	“一带一路”孕灾环境数据知识服务应用	http://drr.ikcest.org/knowledge_service/the_belt_and_road.html	通过网络、文本、统计等多源手段,获得“一带一路”沿线国家基础国情信息,并在线展示和服务
15	“一带一路”中蒙俄经济走廊草地产草量知识应用	http://drr.ikcest.org/knowledge_service/grassland_yield.html	构建中蒙铁路沿线(蒙古段)产草量估算模型,获得长时间序列产品并可视化
16	灾害事件公众舆情分析知识应用	http://drr.ikcest.org/knowledge_service/ncp.html	基于新浪微博大数据,获取中国在新型冠状病毒肺炎疫情期间的公众舆情并可视化

4 新型冠状病毒肺炎灾害舆情分析应用案例

4.1 社交媒体数据处理

(1) 数据获取。通过新浪微博数据中心官方 API,以“冠状病毒”为关键词,获取相关微博文本信息。经空间信息和属性信息提取,形成包括用户名、用户 ID、微博文本、地理位置、发布时间等属性字段的可用信息。

(2) 数据预处理。原始社交媒体数据噪声较大,有很多无效的重复信息,因此要对原始文本内容进行去重去噪、分词、建立索引等处理,使文本数据格式符合模型的输入格式。使用正则表达式对原始社交媒体文本进行过滤,去除干扰信息(例如 http 链接、标点符号)、

停用词、低质量文本、重复文本。中文分词是将句子切分成一个个单独的词的过程。本文使用 Python 中文分词工具“结巴”进行中文分词操作。

4.2 主题抽取与分类框架

基于机器学习算法和主题模型,对中文分词的结果进行话题抽取与分类,构建主题抽取与分类框架^[3]。如图 3 所示,该框架主要由三部分构成:①从预处理后的文本数据选取样本数据,基于 Python 中的“Gensim”库,使用隐含狄利克雷分布主题模型^[4]进行主题抽取,生成各文本的主题概率分布以及各主题的单词概率分布。人工将生成的主题归纳为以下七个一级类:“疫情通报”、“疫情科普”、“官方应对”、“个人应对”、“观点情绪”、“求助信息”、“捐助信息”,对样本数据进行标注。②已标注主题的样本数据被作为随机森林算法^[5]的训练样本,基于 Python 中的“Scikit-learn”库,对整个数据集进行分类。③对每个主题的文本数据应用上述步骤,得到细粒度的二级分类。

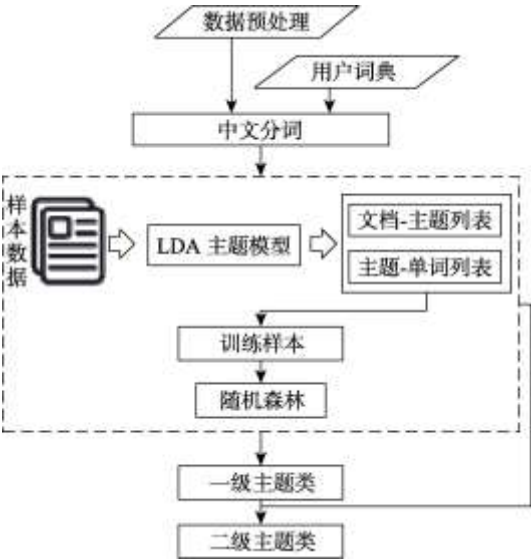


图 3 微博信息主题抽取与分析框架

4.3 时空序列分析应用案例

以 2020 年 1 月 9 日 0 点至 1 月 31 日 24 点的新型冠状病毒肺炎微博文本信息为例,初步获得相关信息 648,013 条,其中具有地理坐标且位于中国区域内有 55,260 条。图 4 是疫情相关微博数量的时间序列分析图。由图 4 可知,1 月 9 日发生在武汉的不明原因肺炎的病原体初步判定为新型冠状病毒后的微博关注量保持平稳和轻微抬升态势,1 月 20 日中央对此次肺炎做出最高指示,同日,钟南山院士肯定新型肺炎存在“人传人”的特点,关注疫情的微博数量开始急剧攀升,在 21 日达到峰值。之后因受春节假期影响关注量振荡下降到 1 月 29 日(正月初五)触底,期间武汉封城以及广东、浙江和湖南等部分省份的响应举措也带来微博信息的显著波动。1 月 31 日 WHO 宣布本次新冠肺炎疫情构成国际关注的突发公共卫生事件,也对微博信息时间轨迹变化带来影响。

图 5 为疫情相关微博的空间分布统计,显示出主要舆情热点集中在湖北、山东、河南、江苏、浙江、四川和广东 7 省。以 200 km 为搜索半径,采用核密度(Kernel Density, KD)分析法,形成疫情舆情关联位置信息的可视化分布图(图 6)。图中显示热点突出集中在以鄂豫交界区、冀鲁交界区、苏浙皖交界区核心热点的三角形高值区域,以及四川、广东两个独立热点地区。

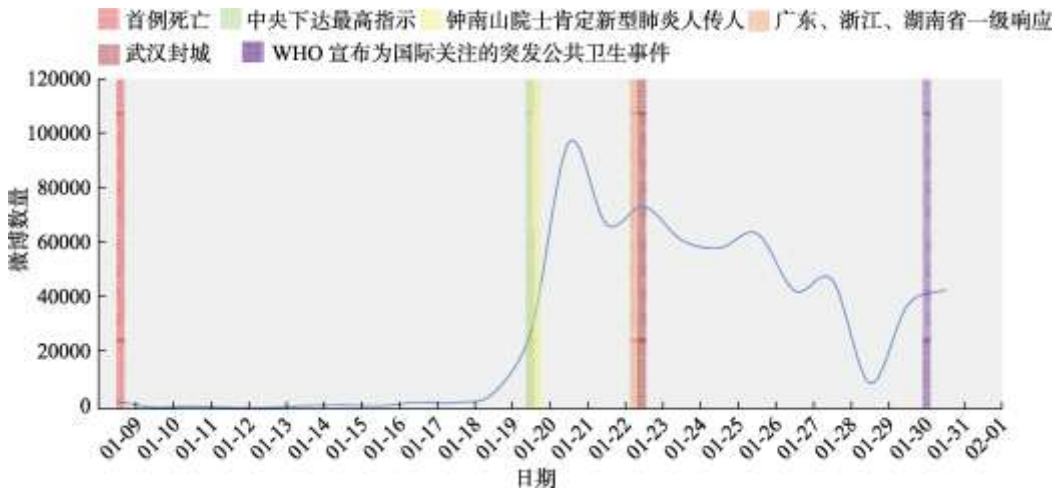


图 4 中国每天与疫情有关的微博数量时间序列

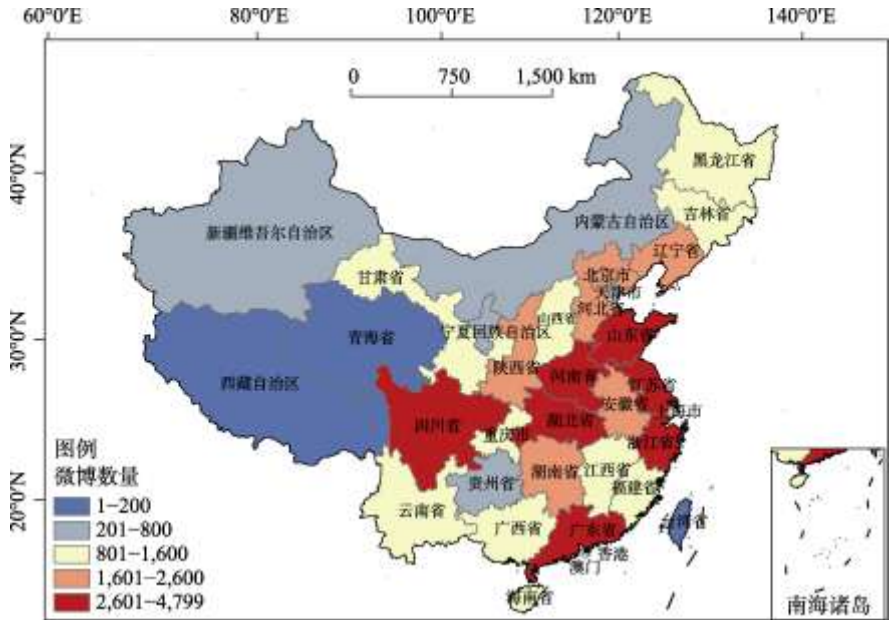


图 5 2020 年 1 月 9 日至 31 日中国疫情相关微博数量空间分布统计（审图号：GS(2019)1831 号）

5 结语

防灾减灾知识服务系统已经实现在线运行服务。截至 2019 年底，防灾减灾知识服务已经为公众提供了 167 个数据集产品、1,050 幅专题地图、9 万个元数据、15 项知识应用、22 万篇文献等服务。在应对当前紧迫的新型冠状病毒肺炎疫情灾害时，防灾减灾知识服务系统快速构建灾害舆情分析在线知识应用，并在知识应用模块进行发布共享和可视化服务。用户可对其进行查询、浏览、下载等，为应急减灾和公共卫生等领域的专家学者进一步分析研究舆情的变化及加强防控举措提供支持。

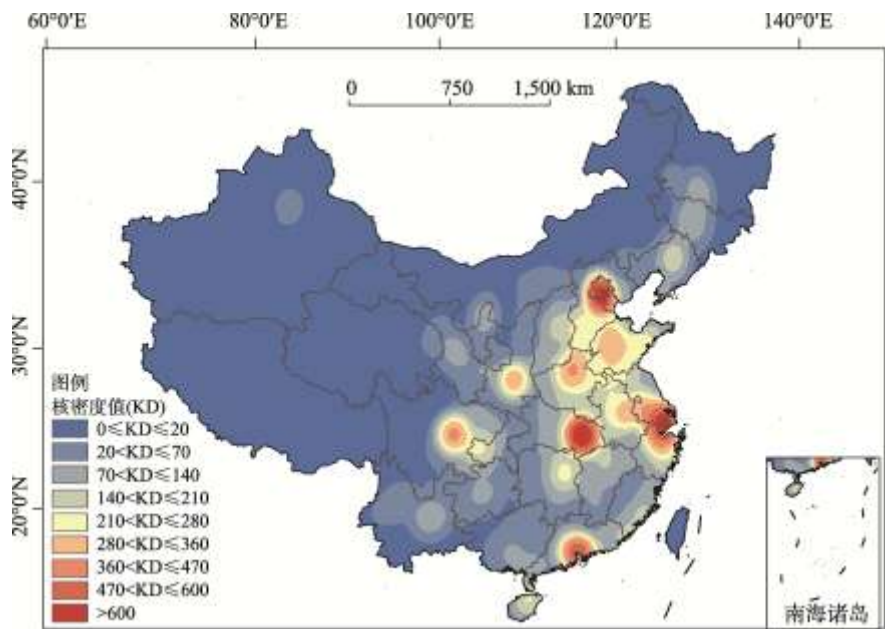


图 6 疫情相关微博核密度图（搜索半径：200 km）（审图号：GS(2019)1831 号）

致谢：感谢防灾减灾知识服务系统的全体参建人员。感谢国际工程科技知识中心秘书处和 UNESCO 防灾减灾部门专家的指导。

参考文献

[1] Wang, J., Bu, K., Yang, F., *et al.* Disaster risk reduction knowledge service: a paradigm shift from disaster data towards knowledge services [J]. *Pure and Applied Geophysics*, 2020, 177(1): 135–148.

[2] 王玉洁, 卜坤, 王卷乐. 基于开源 Pycsw 的灾害元数据管理系统设计与原型实现[J]. 科研信息化技术与应用, 2018, 9(2): 62–72.

[3] Han, X., Wang, J. Using social media to mine and analyze public sentiment during a disaster: a case study of the 2018 Shouguang city flood in China [J]. *ISPRS International Journal of Geo-Information*, 2019, 8(4): 185.

[4] Blei, D. M., Ng, A. Y., Jordan, M. I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3(1): 993–1022.

[5] Breiman, L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5–32.