

Statistical Analysis of Background Data in Internet of Intelligences

Wen Tian, Chongfu Huang

Academy of Disaster Reduction and Emergency Management, Faculty of Geographical Science
Beijing Normal University
Beijing, China
15670994902@163.com

救灾互联网中背景数据统计分析

田雯, 黄崇福

北京师范大学地理科学学部减灾与应急管理研究院
北京 100875, 中国
15670994902@163.com

Abstract—The online computing module supported by geospatial information diffusion technology can obtain complete disaster data through incomplete background data in the early stage of disasters, thus making up the disaster data blank and supporting accurate rescue. Among them, with the advent of the Internet era, the acquisition of background data will realize the transition from traditional static acquisition to dynamic acquisition with networked disaster data as the main body. Systematic background data statistics is the basic work. Therefore, this paper will (1) discuss the concept of background data; (2) take flood disaster as an example, comprehensively consider the rapid assessment indicators of disasters at home and abroad with the theme of disaster relief, and form an online calculation for supporting disaster relief resource matching and disaster relief intelligent network. Background data indicator system and its data collection index items; (3) Statistics and analysis of relevant background data sources of domestic open database and web page data, forming a list of background data sources for collecting data; (4) For example, in Santai County, Sichuan Province, based on the existing database source data items, combined with the MySQL database architecture, the background database design scheme in the disaster relief intelligent network is constructed to provide a data foundation for the online calculation of the disaster relief intelligent network.

Keywords—Internet of Intelligences; background data; flood disaster

摘要—由地理空间信息扩散技术支持的救灾互联网在线计算模块可以在灾害初期通过不完整的背景数据获得完整的灾害数据,从而弥补灾情数据空白、支持精准救援。其中,随着互联网时代的到来,背景数据的获取将实现从传统静态化获取向以网络化灾情数据为主体的动态化获取转变,系统性的背景数据统计是基础性工作。因此,本文将(1)讨论背景数据的概念;(2)以洪涝灾害为例,综合考虑以救灾为主题的国内外灾情快速评估指标,形成用于支持救灾资源匹配、救灾互联网在线计算的背景数据指标体系及其数据采集指标项;(3)对国内开放性数据库和网页数据等多途径的相关背景数据源进行统计和分析,形成用于采集数据的背景数据源列表;(4)以四川省三台县为例,以现有数据库源数据项为根基,结合MySQL数据库体系结构,构建救灾互联网中背景数据库设计方案,为救灾互联网的在线计算提供数据基础。

关键词—互联网, 背景数据, 洪涝灾害

I. 引言

中国受自然灾害影响严重,研究灾情已经成为常态。然而,在灾害初期,由于时空制约,灾情数据往往不完整,数据缺乏使得救援工作的开展出现一段空白期。虽然随着时间的推移,受灾地区的灾情信息会自然地趋于完善,但已经失去救援所需的时效性。如,2008年中国的汶川地震,由于灾情信息获取不及时,灾后数小时后才发现北川是受灾最为严重的区域。

互联网(Internet of Intelligences)这一概念由黄崇福教授在2011年提出^[1],现已发展了包括互联网驱动的风险雷达、海洋环境风险管理互联网平台、风险时效性评价等在内的10个互联网平台^[2],是中国灾害风险分析领域先进的在线灾情评估平台。其中,救灾互联网是由黄崇福教授提出的中国首个以救灾为主题的灾情在线计算平台,其地理空间信息扩散技术支持的互联网在线计算模块可以通过灾害初期的不完整背景数据获得完整的灾情数据,从而为风险管理者做决策提供准确可靠的依据^[3]。

地理空间信息扩散技术依托于背景数据^[4],然而,国内对于背景数据的相关研究较少。因此,通过研究明确背景数据的概念、构建背景数据指标,是统计背景数据的前提工作。为将数据集成,需搭建背景数据库。目前,灾害领域数据库的相关研究已有长足发展,如孙永华等搭建的用于管理洪水灾害空间数据的洪水灾害信息数据库^[5];何萍等搭建的用于评估防灾减灾能力的应急基础数据库^[6];杨斌等搭建的用于评估地震灾害的地震应急基础数据库^[7];等等,但与用于灾情预测的背景数据并不直接相关。因此,考虑到救灾互联网中地理空间信息扩散技术的特殊性,搭建其所需的背景数据库是其智能化计算的前提。救灾互联网平台使用MySQL数据库进行数据管理,因此,为保持与平台的一致性以及数据存储、传递的效率,本文通过MySQL创建数据库、管理数据表。

因此，本文以洪涝灾害为例，基于对背景数据的认识以及背景数据指标的构建，分析其对应的数据采集指标项，进而从国内开放性数据库和网页数据等多途径搜集和统计背景数据源，最终形成面向救灾物联网在线计算的背景数据库。

II. 背景数据与指标构成

A. 背景数据概述

在灾害研究领域，背景数据尚未形成成熟的定义。基于 Tobler 的地理学第一定律，“每一事物都与其他事物有关，与近处事物的关系，比与遥远的事物的关系密切”^[8]，文^[4]给出了背景数据的定义，即存在于被观测单元的具有多个地理特征属性向量的集合，如果已观测数据可成为待观测数据进行估值的媒介，则称其为背景数据集，简称背景数据。例如，用“人口”、“人均 GDP”和“相对暴露度”等数据，依据被观测单元的灾情，我们能够对空白单元的灾情进行估值。此时，“人口”、“人均 GDP”和“相对暴露度”等就是背景数据。一个属性值的数据是否为另一个属性值的背景数据，由两个因素决定，一是能否发挥制约作用；二是能否容易获得。

值得一提的是，背景数据 (Background data) 与基础数据 (Basic data) 这两个概念容易被混淆。经过查证资料总结来看，背景数据更多的是反映事物主体的要素，在时间序列上包括历史和现在，其数据内容相对细致，针对性强；基础一词源于建筑学，强调承上启下的作用，基础数据范围相对小，联系性、连续性强。用于灾害研究的背景数据概念具有专业属性，即具有地理特征属性，包括历史灾情数据、灾害属性数据、自然环境数据、社会经济数据等诸多方面。

B. 背景数据指标

1) 指标选取

救灾物联网中地理空间信息扩散技术的最终目标是构造出“背景数据”和“灾情”之间的因果关系，其构造的前提是收集背景数据 (灾前) 和已观测数据 (灾中)。此外，其主要任务是预测灾区空白单元的灾情，为救灾指挥决策提供参考。据此，我们认为背景数据的组成与灾情指标直接相关，且面向救灾救援。

由于各灾种的背景数据构成不尽相同，因此，本文以气候变化背景下中国最为严重的自然灾害之一——洪涝灾害^[9]为例展开研究。关于洪涝灾害的定义，书^[10]已经给出明确的定义，即当地由于长期降雨或其他原因，使山洪爆发，或使江、河、湖、海所含水体水量迅猛增加，水位急剧上涨超过常规水位时的自然现象，叫做洪水，由此造成的灾害叫做洪水灾害。由于降水过多，地面径流不能及时排除，农田积水超过作物耐淹能力，造成农业减产的危害，叫做涝灾，洪灾和涝灾统称为洪涝灾害。

洪涝灾害是风险源与风险承受体时空耦合的结果，受到多种因素的作用和影响，具有时间和空间变化的特征，因此，背景数据的构成应基于综合性、系统性的灾情评估指标选取。本文通过梳理国内外较为成熟的衡量洪灾灾情的指标，包括国际知名的 HAZUS、DaLA，以及国内现有的反映自然灾害情况和救灾工作情况的《自然灾害情况统计制度》，作为灾情评估指标的选取依据。

2) 指标构成

基于以上分析，将背景数据指标分为社会经济、自然环境、历史灾情、救灾资源四大类，其中，救灾资源和社会经济能够反映灾后居民的自救互救水平。具体指标构成如下表 1 所示：

表 1 背景数据指标构成

类别	指标名称	指标说明	数据采集项
社会经济	人口密度分布	单位面积人口和居民点的密集程度	居民点空间分布
	行政区划	反映该区域人口的易损程度	公里格网人口空间分布
	区域生产总值	反映灾害影响区域和范围	乡镇界线数据集
	耕地分布	区域生产总值反映区域内经济发展程度，水平高低反映易损性程度	公里格网 GDP
	卫生机构	洪水的主要作用对象是耕地	公里格网农业产值人均分布
	生命线设施	灾后医疗资源数量决定救灾质量	土地利用类型
	水利设施	反映灾后居民生活状况	千人床位数
	交通运输	影响洪灾的淹没范围	生命线设施空间分布
	通讯设施	反映交通状况和灾后交通通达性	水利设施空间分布
	自然环境	地形地貌	反映灾后信息传输通达性
历史灾情	河网分布	反映居民位置的相对暴露度	通讯设施空间分布
	历史淹没水深	反映河流与居民的相对位置	DEM 数据集
	历史淹没范围	淹没水深越大，洪涝风险越大	河流水系数据集
	历史淹没历时	淹没范围越广，洪涝风险越大	历史洪峰高度
救灾资源	应急避难场所	淹没时间越长，洪涝风险越大	历史洪灾空间分布
	应急救援队伍	反映受灾人群的可转移率	公里格网降水数据
	应急救援物资	反映救灾的可调配程度	应急避难场所空间分布
	应急救援方案	反映物资与灾民的匹配程度	应急救援队伍数量
		反映实施救灾的具体安排	应急救援物资数量
			应急预案

C. 背景数据源统计

目前，随着互联网时代的到来，数据的网络化使得数据来源多样化。本文主要以国内外使用频率较高的开

放性数据库和网页数据为数据源统计渠道，根据数据采集项分别对应数据源地址，统计结果如下表 2 所示。部分数据需采集后经过一定的处理方可使用。

表 2 背景数据源获取途径

类别	数据源名称	数据获取途径
社会经济	国家数据	http://data.stats.gov.cn/
	统计数据库系统	http://www.chinastatistics.com/
	人地系统主题数据库	http://www.data.ac.cn/index.asp
	中国国家调查数据库	http://cnsda.ruc.edu.cn/index.php
	中国经济与社会发展统计数据库	http://tongji.cnki.net/kns55/Navi/NaviDefault.aspx
	一带一路统计数据库	http://ydyl.cei.cn/page/Default.aspx
	欧盟统计局数据库	https://ec.europa.eu/eurostat/data/database
	世界银行公开数据	https://data.worldbank.org.cn/
	地理国情监测云平台	http://www.dsac.cn/
	地理空间数据云	http://www.gscloud.cn/
自然环境	USGS	https://www.usgs.gov/
	中科院数据云	http://www.csdb.cn/
	国家卫星气象中心	http://satellite.nsmc.org.cn/portalsite/default.aspx
	NASA earth observation	https://neo.sci.gsfc.nasa.gov/
历史灾情	中国气象数据网	http://data.cma.cn/
救灾资源	地方政府官网 实地调研获取	

III. 背景数据库构建方案

A. 背景数据采集

本文以四川省三台县为例构建其洪涝灾害背景数据库，原因有两点：四川省三台县曾于 2018 年 7 月 11 日发生特大洪灾，该事件可作为典型的洪涝灾害研究对象；四川三台县为救灾互联网的示范区，实地获取数据较为便捷。根据以上背景数据指标展开数据采集工作：

- (1) 居民点空间分布、公里格网人口空间分布通过全国人口密度数据产品剪裁获取，为地图数据；
- (2) 由于乡镇级别的边界数据具有一定的密级，需经过购买获取，仅供研究使用，为地图数据；
- (3) 公里格网 GDP 通过全国 GDP 公里格网数据产品剪裁获取，为地图数据；
- (4) 公里格网农业产值人均分布通过栅格代数计算农作物种植面积与作物平均产值以及人口分布的关系得到，为地图数据；
- (5) 土地利用类型通过下载土地利用数据产品剪裁得到，为地图数据；
- (6) 千人床位数通过医疗资源统计数据与人口密度的关系综合得到，为表格数据；
- (7) 生命线设施空间分布、水利设施空间分布、通讯设施空间分布通过实地调研获得，需进行数字化，为地图数据；
- (8) 公里格网路网数据、DEM 数据集、河流水系数据集通过下载路网数据产品以及全国三级河流数据剪裁获得，为地图数据；
- (9) 历史洪峰高度通过三台县政府官网的相关新闻得到，为表格数据；
- (10) 历史洪灾空间分布通过相关信息获取并进行数字化得到，其依据是公里格网降水数

据，即最大三日降雨量超过洪水灾害临界值则记为洪灾，为地图数据；(11) 应急避难场所空间分布、应急救援队伍数量、应急救灾物资数量、应急预案均通过实地调研获取，为表格数据。在采集过程中，部分数据需进行数字化处理，如水利设施空间分布需通过实地获取数据后在地图上标注进行数字化。

B. 背景数据特征

以上所采集数据包括地图数据和表格数据两种。其中，空间数据采用的是矢量或栅格数据结构，而属性数据、规则数据则采用了关系数据模型。值得一提的是，数据采集过程中遇到以下几个与数据特征相关的问题：

- (1) 数据的分辨率问题。由于数据下载路径的不同，数据之间存在分辨率不一的问题，容易出现误差问题；
- (2) 数据产品的统计时间。数据下载中发现数据的时间维度不一，部分数据较老；
- (3) 不同数据来源的数据接口问题；等等。以上问题可能导致在计算的过程中增加一定的误差。

C. 互联网中的背景数据库

1) 数据表设计

目前，MySQL 数据库通过空间数据类型 Geometry 存储，通过几何对象来表示位置，即通过建立位置信息表，每条信息对应的经纬度都会以 Geometry 类型存在表中。然而，将几何数据存储为 BLOB（二进制大对象）数据，在空间数据上只能创建前缀索引，当涉及空间搜索时非常低效，尤其是在涉及复杂的几何数据时。据此，本文选用属性表格式存储数据，即通过 ArcGIS 把地图的属性表导入到 SQL 数据库中，同时将数据处理为

具有经纬度位置属性的属性表数据格式，具体的特征属性值需进行归一化处理。由于地图数量较多，在此只对出现频率较多的字段进行说明，如表 3 所示：

表 3 背景数据表

字段	字段说明	字段类型
ID	标识	varchar(30)
Name	乡镇名称	varchar(100)
Lon	纬度	double(20)
Lat	经度	double(20)
data	特征属性值	double(20)

2) 数据管理

文^[11]已经对救灾物联网中地理信息扩散技术的数据库系统及数据管理进行介绍，即通过使用 PHP 语言编写 SQL 原生语句实现对 MySQL 数据库进行数据查找、增加、删除、更新、导出、恢复等操作，或采用 Yii 框架中的 db 组件来进行数据管理。其对背景数据表具有一定的要求，即需以研究区行政单元为单位进行记录。本文中基于 MySQL 所构建的背景数据库中的数据表具有经纬度信息，可通过 PHP 实时调用，便于实现数据的可视化，其中，各个数据表中的特征属性值字段为地理信息扩散技术可用的计算指标。

IV. 结论与讨论

应地理空间信息扩散技术中背景数据的采集要求，建立背景数据量化指标体系，确定和规范背景数据指标的采集项目和格式，构建符合物联网系统要求的背景数据统计系统，并建立相应的数据库为在线计算提供数据基础。以四川省三台县为例，本文设计了背景数据库设计方案，并以区域内洪涝灾害为研究对象，结合已统计数据源针对性地进行数据采集和存储，研究发现通过导出具有经纬度特征的属性表作为背景数据值能够提高计算效率，此外，数据分辨率、数据产品发布时间等可能进一步造成计算误差。对于数据质量问题，通过数据质量评价模型进行判断是较好选择，此外，尽可能多的寻找数据采集路径是亟待解决的问题。区别于一般的灾情评估方法，地理信息扩散技术所需已知观测数据是灾害初期可获得的数据，其数据采集的指标根据灾后救援所需参考的灾情指标、以及灾后初期通过网络和上报数据可行的数据指标确定。灾害初期的上报数据、网络数据往往会有部分区域的真实数据，而这些数据不仅具有一定的可信度，而且还具有相当程度上的时效性，因此，通

过这些局部已知观测区域的灾情数据按照背景数据扩散出未知地理单元中的灾情数据，即是救灾物联网中的地理空间信息扩散技术的功能所在。对于已观测数据，即实时灾情数据的采集将在下一步研究中展开探索。

致谢

本项目受国家重点研发计划课题“重大自然灾害多层次精准救助关键技术研究”（编号：2017YFC1502902）资助。

参考文献

- [1] 黄崇福. “风险分析在线服务的物联网,” *Journal of Risk Analysis and Crisis Response*, vol. 1, issue 2, pp. 110-117, 2011.
- [2] 黄崇福. “物联网原理及其核心技术的发展,” *Journal of Risk Analysis and Crisis Response*, vol. 7, issue 3, pp. 146 - 155, 2017.
- [3] 黄崇福. “地理空间上的信息扩散及其在风险分析中的应用,” *Journal of Risk Analysis and Crisis Response*, vol. 66, 2018.
- [4] 黄崇福. “借助背景数据的地理空间信息扩散技术,” *Journal of Risk Analysis and Crisis Response*, vol. 9, issue 1, pp. 2-10, 2019.
- [5] 孙永华, 李小娟, 尹连旺, et al. “基于 ArcSDE 和 SQLServer2000 洪水灾害信息数据库的设计研究,” *河北师范大学学报(自然科学版)*, vol. 31, issue 3, pp. 400-402, 2007.
- [6] 何萍, 陈晓发, 傅冠华, et al. 2010. 基于应急基础数据库的城市防震减灾能力评价指标研究[J]. *华南地震*, 30(2):36-45.
- [7] 杨斌, 徐博, 高翔. “地震应急基础数据检查系统的设计与实现,” *震灾防御技术*, vol. 11, issue 1, 2016.
- [8] Tobler W R. “A computer movie simulating urban growth in the Detroit region,” *Economic Geography*, vol. 46, pp. 234-240, 1970.
- [9] 刘志雨, 夏军. “气候变化对中国洪涝灾害风险的影响,” *自然杂志*, vol. 38, issue 3, pp. 182-188, 2016.
- [10] 黄崇福. *自然灾害风险分析与管理*. 北京: 科学出版社, 2012.
- [11] Yiran Yin, Chongfu Huang. “A Research on Coding Information Diffusion on Geographical Space in IOI for Disaster Mitigation,” *Journal of Risk Analysis and Crisis Response*, vol. 66, 2018.