

引用格式:刘业森,张晓蕾,郭良.自然灾害调查数据的多尺度异常检测方法研究及应用[J].地球信息科学学报,2017,19(12):1653-1660. [Liu Y S, Zhang X L, Guo L. 2017. Study and application of the method of multi-scale outliers detection of natural disaster investigation data. Journal of Geo-information Science, 19(12):1653-1660.] DOI:10.3724/SP.J.1047.2017.01653

自然灾害调查数据的多尺度异常检测方法研究及应用

刘业森¹, 张晓蕾^{2,3}, 郭良^{2,3*}

1. 天津大学 水利工程仿真与安全国家重点实验室, 天津 300072; 2. 中国水利水电科学研究院, 北京 100038;
3. 水利部防洪抗旱减灾工程技术研究中心, 北京 100038

Study and Application of the Method of Multi-scale Outliers Detection of Natural Disaster Investigation Data

LIU Yesen¹, ZHANG Xiaolei^{2,3} and GUO Liang^{2,3*}

1. State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China; 2. China Institute of Water Resources and Hydropower Research, Beijing 100038, China; 3. Research Center on Flood & Drought Disaster Reduction of the Ministry of Water Resources, Beijing 100038, China

Abstract: "Natural disaster" is the phenomenon of the losses of life and property, which is caused by the interaction of human society and natural environment. It's also the product of the disaster environment, disaster-causing factors and disaster-bearing body. In order to study the processes, mechanisms and impacts of natural disasters as well as the reduction of the losses caused by natural disasters, it is necessary to conduct surveys of basic data and natural disaster events on a large scale of which the authenticity and consistency are much significant for ensuring the reliability and validity of the research results. However, the large number of organizations and investigators participating in the survey and evaluation process, large regional differences and large spatial scale create challenges in data quality control and validating the consistency of data from various survey units. To ensure the correctness and consistency of the data, it is necessary to carry out manual inspection. However, for the massive survey data, it is unrealistic to totally rely on manual work to effectively identify the abnormalities. As a result, we design a multi-scale anomaly detection method for natural disaster survey data by using the single-element detection method of outliers based on normal distribution and spatial clustering method of Anselin Local Moran's I to detect the abnormal values and abnormal spatial distribution patterns of the massive survey data. It can effectively extracts the abnormalities and abnormal investigation units at all levels of scale and gains the reasons for abnormal data. It provides the support for the manual checking of survey data. In this paper, taking the project of flash flood disaster investigation and evaluation in mainland of China as an example, this method is used to audit the events of historical flash flood disaster and the areas of the towns which are in the prevention zones. Also, it quickly extract the anomaly units of flash flood disaster point density and township units with exceptional area values. Further analysis found that the reasons for these abnormalities were due to the inconsistency of filling methods, unit errors, and repetition of records and so on. The method resolved the inconsistency in massive amounts of flash flood survey data. This method is an effective approach

收稿日期 2017-07-10; 修回日期: 2017-09-06.

基金项目: 中国水科院科研专项(JZ0145B042016、JZ0145C022017); 国家自然科学基金项目(51579131)。

作者简介: 刘业森(1980-), 男, 博士生, 研究方向为遥感、GIS及其在水利行业的应用。E-mail: yesenl@lreis.ac.cn

*通讯作者 郭良(1960-), 教高, 研究领方向为山洪灾害防治。E-mail: guol@iwhr.com

of checking the quality of various other large-scale disaster datasets. Although the data validation approach used in this study is very effective, there are still some problems, i.e. the outlier checking only considers the outliers between survey units based on the administrative divisions. Regions are not divided according to their economic development and natural conditions. Finally, we analyze the applicable conditions of this method in the large-scale natural disaster investigations.

Key words: disaster investigation; flash flood disasters; data quality; anomaly detection; spatial clustering

***Corresponding author:** GUO Liang, E-mail: guol@iwhr.com

摘要: 大范围自然灾害调查,涉及区域环境差异大,数据获取方式多样,参与人员多,各级汇总成果中存在一些异常调查单元,需要人工判读其合理性,但单纯依靠人工从海量数据中有效识别异常是不现实的。本文设计了一种自然灾害调查数据的多尺度异常检测方法,综合运用离群检测方法和空间数据挖掘算法,分别进行异常值和异常空间分布模式检测,能够从海量调查数据中快速提取各级尺度的异常值和异常调查单元,支撑人工判读工作。将该方法应用于全国山洪灾害调查评价汇总数据的审核中,以全国历史山洪灾害点和防治区乡镇面积审核为例,分别快速提取了县乡两级区划中的山洪灾害点密度异常单元和面积值异常的乡镇单元,通过对检测结果进行分析,发现是填报口径不一致、单位错误、记录重复等原因造成的。最后分析了该方法在大范围自然灾害调查中的适用条件和方法。

关键词: 灾害调查;山洪灾害;数据质量;异常检测;空间聚类

1 引言

“自然灾害”是人类赖以生存的自然界作用于人类社会并造成生命财产损失的现象与事件,是孕灾环境、致灾因子、承灾体相互作用的结果^[1]。为了避免或减少灾害损失,需在大量调查数据的基础上,研究灾害发生的规律、过程、机理、影响等。灾害调查数据的真实性和一致性是保证研究结果可靠有效的前提和关键因素。自然灾害具有突发性、分布广、发生频繁、随机性强的特点,具有多维联系性和尺度效应^[2-4],大范围自然灾害调查涉及环境条件复杂、数据值域难以确定,大部分数据项并无明确的控制规则;调查工作需要多人参与,调查人员认知水平不一必然会导致大范围调查数据中存在区域一致性问题。另外,随着数据采集与测量技术的发展,灾害调查方法在传统的调查问卷、历史资料收集等方法的基础上,增加了物联网、互联网、移动测量等新技术的应用,不同技术手段获取的数据,数据特征、深度会有差别。大范围自然灾害调查需要从基层调查单元层层汇总,形成最终的数据集,由于调查方法和调查人员方面的因素,会导致各级汇总成果中存在一些异常调查单元,需要人工判读其合理性,而单纯依靠人工从海量数据中有效识别异常单元是不现实的。因此,需要一种有效的技术支撑手段,能从大范围自然灾害调查数据集中快速发现异常数据^[5-6]。

传统的调查数据质量保证方法,大多从数据生产端进行数据质量控制,利用指标值域、基数控制、拓扑关系等,检查数值精度,达到数据质量控制目的^[7-9],对于具有复杂逻辑关系和拓扑关系的数据,则利用数据模型,维护数据对象关系和数据精度^[10-11],这些方法能奏效的前提是数据具有明确的规则,但自然灾害发生环境空间异质性决定了灾害调查数据存在明显的区域差异,且由于调查手段、调查口径、工作组织等方面的原因,存在一些无法预期的异常问题,因而无法进行预先的规则设置。对于海量数据异常值或异常模式的检测,在互联网、电商、电信、工业制造等领域的运营数据或实时监测数据的质量控制中应用较为广泛,大多采用计量公式、聚类、数据挖掘等方法进行异常值检测^[12-16]。具有时空属性的数据,异常检测方法相对复杂,需要综合运用统计方法、空间聚类、时空数据挖掘等方法^[17-18]。在已开展的大范围调查统计实践中,灵活采用了不同的异常检测方法。在全国经济普查、全国人口普查、全国土地调查、全国水利普查、全国环境统计等统计调查工作中,主要采用完整性检查、逻辑关系审核、历史数据比对、数据范式、值域、拓扑关系、数据模型等方法,保证数据质量^[19-21]。大范围自然灾害调查数据中异常的不确定性和多级综合效应,决定了在异常数据检测过程中,必须综合考虑各级调查单元的异常问题。目前的异常数据检测方法,不能完全满足大范围自然灾害调查数据的异常检测。

本文通过建立一种基于离群点检测和空间聚类方法的异常检测方法。综合运用统计学中的基于正态分布的离群点一元检测统计方法和空间聚类中的 Anselin Local Moran's I 检测调查数据中的异常属性值和异常数据单元。通过对异常数据的多尺度下钻,能从海量数据中快速发现异常数据及查找异常数据的来源。

2 异常检测方法设计

2.1 异常值检测

大范围自然灾害调查数据的数值型数据包括面积、高程、人口等属性数据,以及基于调查单元(行政区划、流域、自然分区、经济分区、农业分区等)汇总的数量和密度数据。致灾因素突发异常的特点和自然条件的空间异质性,会造成调查数据中形成正常的离群点,同时,调查工作偏差也会形成离群点,这就需要人工判读这些离群点是否合理。人工判读的基础是能够从海量数据中快速检测出离群点,从而进行专业确认或人工审核。采用统计学中的基于正态分布的离群点一元检测统计方法,可以快速提取离群点,处理方法是利用数据项的绝对值或标准差进行检测^[22-23]。

(1)数据项与全部数据项统计的四分位数 $Q3$ 或 $Q1$ 差值的绝对值大于 1.5 倍四分位数极差,检测为离群点: $x-Q3>1.5\times IQR$, $Q1-x>1.5\times IQR$ (IQR 为四分位数 $Q3$ -四分位数 $Q1$)。

(2)数据项减去所有数据项平均值的绝对值大于 3 倍标准差(分布在 6σ 外)为离群点:即 $|x-\bar{x}|>3\sigma$ 。不符合正态分布的数据,则通过排序选取数值最高和最低的 1% 的数据项作为离群点。通过离群点检测方法,可从海量调查数据中快速筛选异常值。

2.2 异常空间分布模式检测

空间分布模式包括聚集、离散、负相关等,不同的分布模式体现不同的分布特征^[24]。自然灾害受气候、植被、地形、高程、人口分布等诸多因素影响,空间分布模式复杂多样,不同灾种发生机理不同,空间分布模式也不同。调查人员对灾害样本的主观判断差异,以及调查方法、行为习惯差异,都会造成不合理的异常空间分布模式,这些分布模式存在于各级调查单元中。异常空间分布模式检测的目的是从海量数据中快速自动提取这些异常分布模

式。根据地理学第一定律^[25],相近的单元更相似,在灾害调查中,异常检测所基于的原则是如果某一调查单元与周边单元具有相似的自然条件和社会经济条件,则调查结果应该具有一定的相似性。本文利用 Anselin Local Moran's I 方法来识别数据对象的空间分布模式是否异常^[26]。该方法的输出结果为 Local Moran's I 指数、 z 得分、 p 值和聚类/异常值类型。 z 得分和 p 值用来度量统计量的显著性,判断输入对象的相似性或相异性。如果要素的 z 得分是一个较高的正值,则表示输入对象与周围的对象相似;如果要素的 z 得分是一个较低的负值,则表示有一个具有统计显著性的异常对象,输出的结果中将显示该调查单元是低值中间的高值,亦或是高值中间的低值。

调查单元的 Local Moran's I 计算方法见式(1)。

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{ij}(x_j - \bar{X}) \quad (1)$$

式中: x_i 是第 i 个调查对象或调查单元的指标值; \bar{X} 是所有调查对象或调查单元相应指标的平均值; w_{ij} 是对象 i 和 j 之间的空间权重。 S_i^2 的计算方法见式(2)。

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{X})^2}{n-1} - \bar{X}^2 \quad (2)$$

式中: n 等于调查对象或调查单元的总数。

z_{I_i} 得分的计算方法见式(3)、(4)。

$$z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \quad (3)$$

其中:

$$E[I_i] = -\frac{\sum_{j=1, j \neq i}^n w_{ij}}{n-1} \quad (4)$$

$$V[I_i] = E[I_i^2] - E[I_i]^2 \quad (5)$$

通过 Local Moran's I , 可以判断任一调查单元与周边调查单元的空间分布模式,从而可以快速地筛选出异常调查单元。

2.3 多尺度异常检测方法

全国或大范围自然灾害调查需要多人参与,层层汇总,形成最终数据集,因此,需要在各层级保证数据的总量、密度、比值等指标的合理性。首先,检测各级汇总指标中的异常值,如人口、面积、高程、财产等数值类信息是否存在明显的离群值,在确认

离群值合理的基础上,进行异常空间分布模式检测。从大尺度到小尺度逐级检测,对异常调查单元进行层层下钻,最小粒度到记录层级,可以采用行政区划的省、县、乡、村作为调查单元序列,也可以采用流域的逐级汇流关系作为调查单元序列,如图1所示。通过逐层提取异常调查单元,作为人工判读的基础。

本文设计的异常检测方法,从数据属性值、空间分布两个维度进行异常检测,从不同尺度保证数据质量。该方法的主要目的是解决海量数据人工审核工作量大的问题,为人工审核提供支撑。

3 异常检测方法在全国山洪灾害调查评价中的应用

3.1 全国山洪灾害调查评价概况

山洪灾害是目前中国造成人员伤亡的主要自然灾害种之一^[27-28]。为了更好地开展山洪灾害防治工作,水利部、财政部启动了山洪灾害调查评价工作,在全国30个省(市、自治区)、305个市、2058个县开展山洪灾害调查评价^[29-30]。调查数据最小尺度到自然村,包括自然村人口、面积、房屋、村貌照片、沿河村落住户宅基地高程、河道断面、沿河村落预警指标

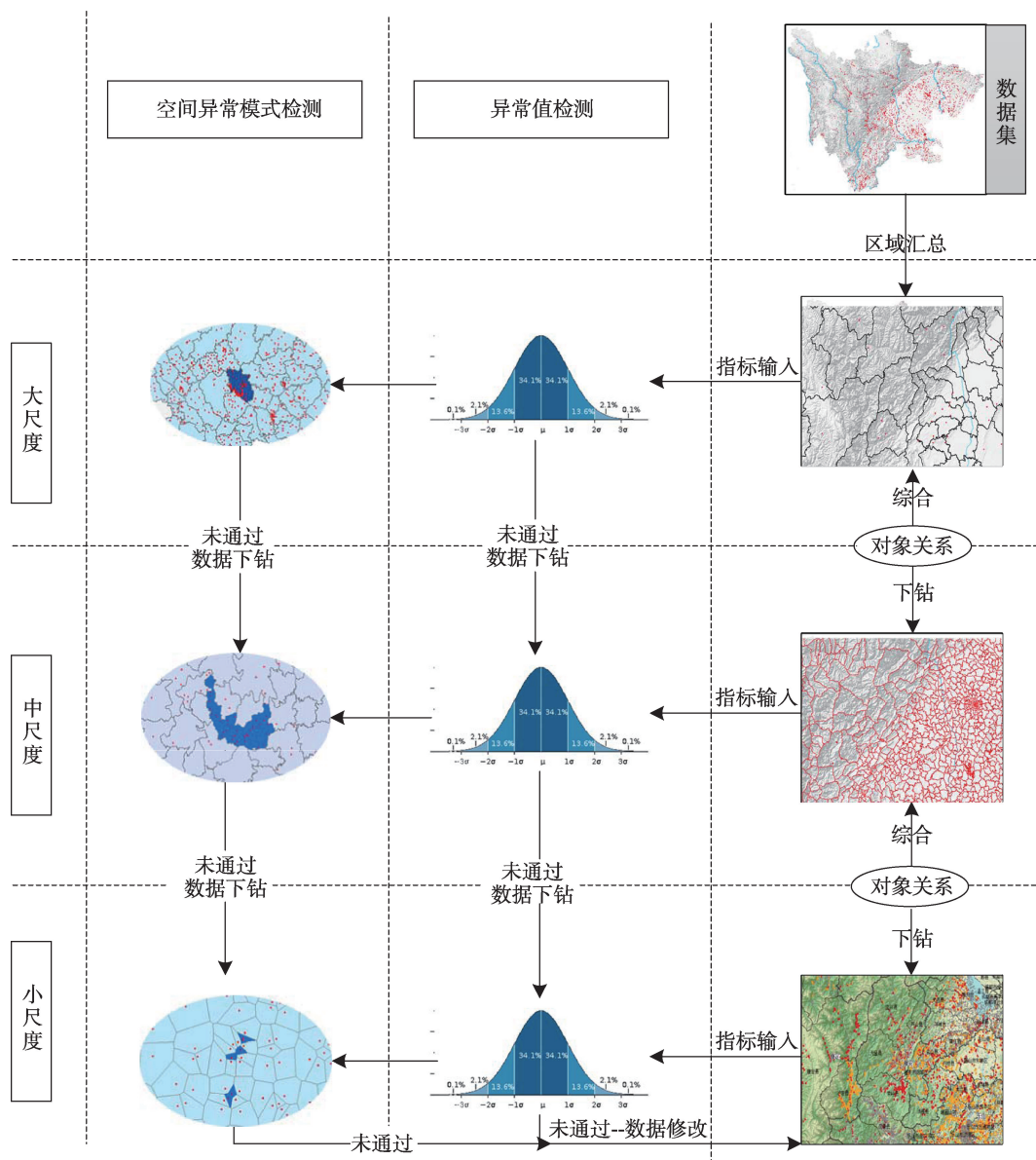


图1 异常检测流程

Fig. 1 Flowchart of the abnormal data detection

等,涉及省、市、县、乡、村、组户的两百多万个行政区划,参与调查的单位600余家,调查人员12万余人。各级调查评价数据经过多批次逐层级汇总,形成全国汇总数据集。为保证数据质量,需要对每批次数据进行审核,审核工作量巨大,并要兼顾数据审核精度和效率。本文设计的异常检测方法,可以从“绝对”(异常值)和“相对”(异常分布模式)2个角度发现数据中存在的问题,所采用的统计方法和空间聚类方法,能从海量数据中快速提取可能的异常数据,为人工判读数据质量和合理性提供支撑。

3.2 检测指标设计

结合全国山洪灾害调查评价数据范围广、数据量大、层级多等特点,根据全国山洪灾害调查评价数据质量审核要求,为各级调查单元分别构建相应的数据审核指标,作为异常值和异常空间分布模式检测的基础(表1)。

3.3 异常检测结果

以历史山洪灾害事件和防治区乡镇土地面积2个指标项的检测为例进行说明。历史山洪灾害点为汇总型数据,要保证数据分布合理;防治区乡镇土地面积为乡镇对象的属性项,要保证属性值在合理范围内。

(1)历史山洪灾害事件

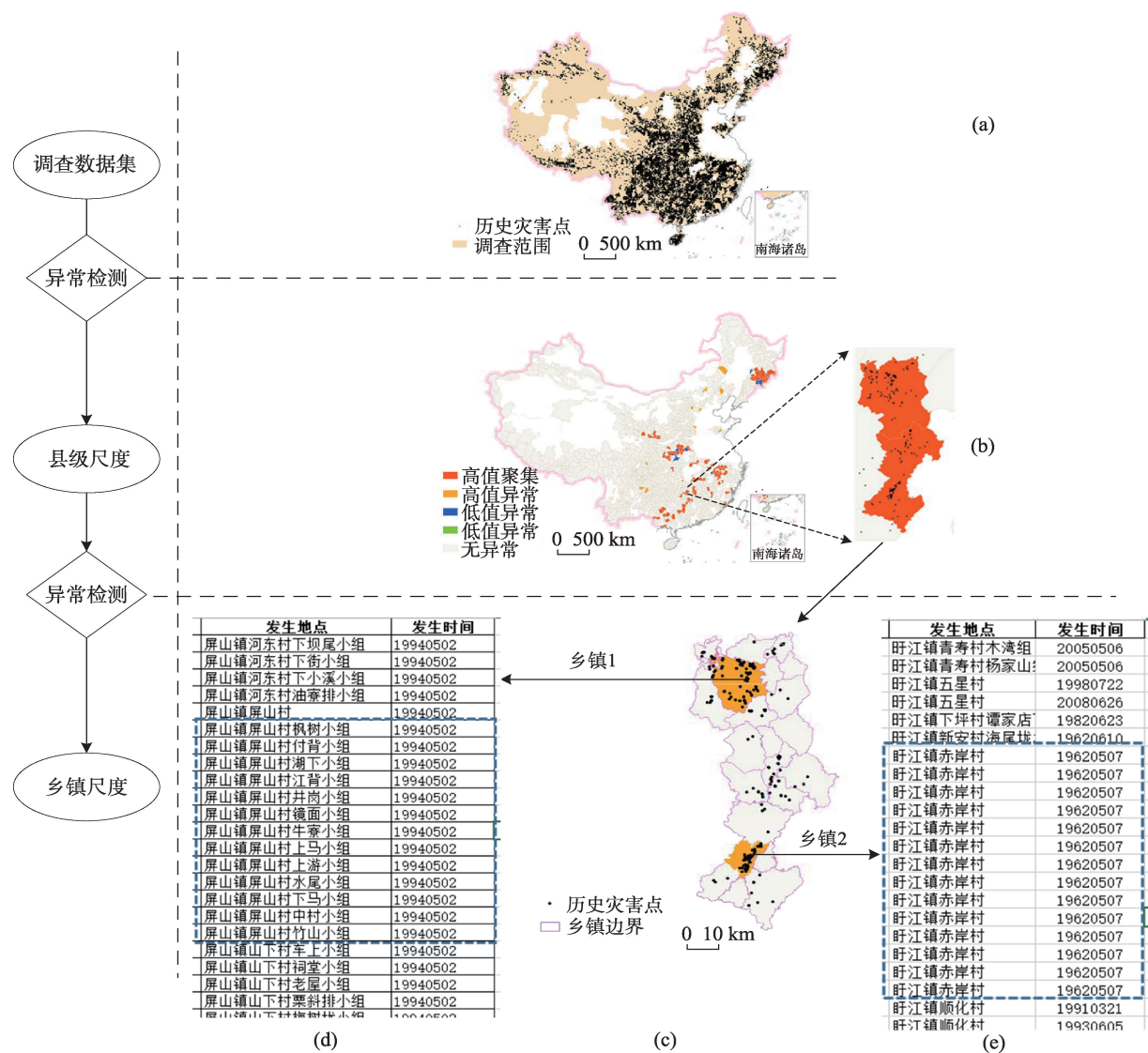
按全国山洪灾害调查技术要求,历史山洪灾害调查以县级行政区划为工作单元,每个防治县均需调查建国以来发生的历次山洪灾害,图2(a)为某时点汇总的全国59 256个历史山洪灾害点。该批次数据包括1660个县级单元,空间分布异常检测结果发现,104个县高值聚集,11个县高值异常,13个县低值异常,1532个县无异常,分布情况见图2(b)。对异常县级单元下钻到乡级尺度进行异常分布检测,图2(c)为其中2个相邻县级单元的检测结果,26个乡镇中有2个乡镇出现高值异常。为了确认结果准确性,查看数据属性表,发现乡镇1出现异常的原因是,将一次山洪过程按村组填写,填报的37条记录中,包括8个村落,其他29条都是这8个村落内的村组(图2(d)),乡镇2出现异常的原因是一次山洪事件重复填写,47条记录中有24条重复(图2(e))。

(2)防治区乡镇土地面积

对某批次数据中29 502个乡镇的土地面积统计结果显示(图3),该数据不符合正态分布特征,因此采用极值百分比方式,选取面积最大的1%的乡镇。最大值1%乡镇和Anselin Local Moran's *I*异常检测结果见图4,其中1%面积最大的乡镇数为295个,Anselin Local Moran's *I*检测异常乡镇数为

表1 山洪灾害调查数据异常数据检测指标
Tab. 1 Detection index of flood disaster investigation data

| 数据尺度 | 调查对象/类别 | 审核指标 |
|------|---------|--|
| 村级 | 防治村 | 人口、户数、土地面积、耕地面积 |
| | 危险区 | 人口、总户数、各类经济户数、各类房屋数 |
| | 监测预警设施 | 自动监测站的集水面积 |
| | 山洪沟 | 集水面积、长度、比降、影响耕地、影响人口等 |
| | 涉水工程 | 塘堰坝:容积、坝高、坝长;路涵:涵洞高、长、宽;桥梁:桥长、宽、高 |
| | 历史山洪 | 过程降雨量、死亡人数、经济损失、损毁房屋等 |
| | 洪水调查 | 最高水位、各时段雨强、洪水流量 |
| | 断面测量 | 历史最高水位、糙率、高程等 |
| 乡镇级 | 数量 | 各类工程总数、历史山洪灾害总数、企事业单位总数等 |
| | 比值 | 单位河流长度各类工程数量、企事业单位数量/县域内人口数量、监测预警设施数量与各县非工程措施建设成果、各乡镇人口与户数比值 |
| | 密度 | 各类工程密度、危险区内单位面积居民户数量 |
| 县级 | 数量 | 同乡镇级 |
| | 比值 | 同乡镇级;各经济指标、人口与2014年统计年鉴对比 |
| | 密度 | 同乡镇级 |
| 省级 | 数量 | 同县级 |
| | 比值 | 同县级;各类调查对象与2013–2015年实施方案对比 |
| | 密度 | 同县级(以省内山丘区面积作为分母) |



注:图(a)为全国历史山洪灾害点分布;图(b)为全国历史山洪灾害点数据县级异常检测结果;图(c)为数据异常县的乡镇级异常检测结果;图(d)为异常乡镇1的灾害点属性表;图(e)为异常乡镇2的灾害点属性表

图2 历史山洪灾害点数据审核结果

Fig. 2 Auditing results of the point data of historical flash flood disaster

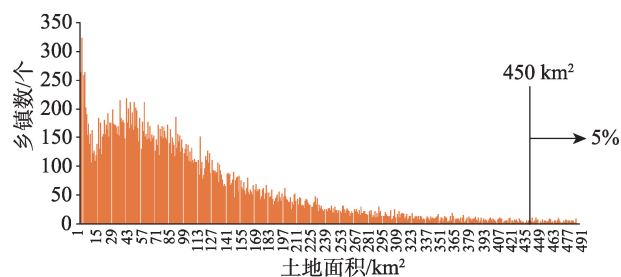


图3 乡镇土地面积统计结果

Fig. 3 land area of township

121个。检查确认发现,1%最大的295个乡镇,有82个为错误数据,其他213个数据无问题。Anselin

Local Moran's I 检测异常的121个乡镇,其中3个乡镇数据无问题,其他118个乡镇数据填写错误。

异常值检测的依据为属性值域,如果区域之间实际情况差异过大,会将正常属性值检测为异常值,如新疆、西藏等西部省份的部分乡镇面积是东部省份乡镇面积的几十倍,这种情况下就会将正常的面积值判断为异常值。而异常空间分布模式的检测,以空间单元与周边单元的关系为检测依据。在全国山洪灾害调查评价汇总数据审核中,综合运用2种异常检测方法,可有效发现数据中存在的问题。

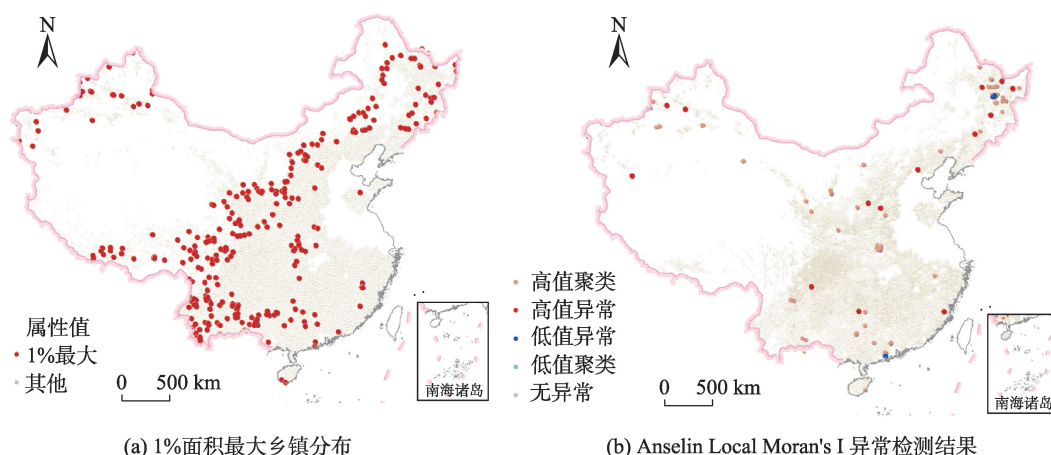


图4 乡镇土地面积异常检测结果

Fig. 4 Spatial distribution of outliers of the rural land area

4 结论

本文建立了一种基于离群点检测和空间聚类方法的异常检测方法。利用统计学中的基于正态分布的离群点一元检测统计方法,检测海量数据中的异常属性值;利用Anselin Local Moran's I 检测调查区域内的异常数据单元。可对海量调查数据的异常属性值和分布模式进行快速检测。

在全国山洪灾害调查评价数据审核中,运用本文建立的异常数据检测方法,有效地解决了海量数据中的异常数据的快速检测。以历史山洪灾害事件调查数据和乡镇土地面积2类数据为例,从全国汇总数据中自动提取了县、乡两级的异常调查单元,对异常检测结果的分析发现造成这种异常的原因包括,数据单位错误、统计口径差异、数据重复等。

本文的方法主要适用于大范围自然灾害类调查数据的异常检测,需要根据灾种特点构建合理的检测指标。另外,异常检测只是保证数据质量控制的一方面,在具体调查实践中,技术要求、调查人员培训、事前事中的过程控制、数据阶段审核等,都是保证数据质量的重要措施。

参考文献(References):

- [1] 史培军,吕丽莉,汪明等.灾害系统:灾害群、灾害链、灾害遭遇[J].自然灾害学报,2014,23(6):1-12. [Shi P J, Lv L L, Wang M, et al. Disaster system: Disaster cluster, disaster chain and disaster compound[J]. Journal of Natural Disasters, 2014,23(6):1-12.]
- [2] 黄崇福.自然灾害基本定义的探讨[J].自然灾害学报,2009,18(5):41-50. [Huang C F. A discussion on basic def-

inition of natural disaster[J]. Journal of Natural Disasters, 2009,18(5):41-50.]

- [3] 赵思健.自然灾害风险分析的时空尺度初探[J].灾害学,2012,27(2):1-6. [Zhao S J. A preliminary study on the spatial and temporal scales of natural disaster risk analysis [J]. Journal of Catastrophology, 2012,27(2):1-6.]
- [4] 刘毅,杨宇.历史时期中国重大自然灾害时空分异特征[J].地理学报,2012,67(3):291-300. [Liu Y, Yang Y. Spatial distribution of major natural disasters of China in historical period[J]. Acta Geographica Sinica, 2012,67(3):291-300.]
- [5] 王宏志.大数据质量管理:问题与研究进展[J].科技导报,2014,32(34):78-84. [Wang H Z. Big data quality management: problems and progress[J]. Science & Technology Review, 2014,32(34):78-84.]
- [6] 李永红,范立民,贺卫中,等.对如何做好地质灾害详细调查工作的探讨[J].灾害学,2016,31(1):102-112. [Li Y H, Fan L M, He W Z, et al. Discussion on how to do the detailed investigation of geological hazards better[J]. Journal of Catastrophology, 2016,31(1):102-112.]
- [7] Morton M, Levy J L. Challenges in disaster data collection during recent disasters[J]. Prehospital & Disaster Medicine, 2011,26(3):196-201.
- [8] 曾五一.国家统计数据质量研究的基本问题[J].商业经济与管理,2010,1(12):72-76. [Zeng W Y. The research on the basic issues of the statistical data quality of the government[J]. Journal of Business Economics, 2010,1(12):72-76.]
- [9] 韩京宇,徐立臻,董逸生.数据质量研究综述[J].计算机科学,2008,35(2):1-5. [Han J Y, Xu L Z, Dong Y S. An overview of data quality research[J]. Computer Science, 2008,35(2):1-5.]

- [10] 程益联,郭悦.水利普查数据质量控制的研究[J].水利信息化,2012(3):1-4. [Cheng Y L, Guo Y. Research on data quality control of national census for water [J]. Water Resources Information, 2012,3:1-4.]
- [11] Berrahou L, Lalande N, Serrano E, et al. A quality-aware spatial data warehouse for querying hydroecological data [J]. Computers & Geosciences, 2015,85(PA):126-135.
- [12] 段华明,何阳.大数据对于灾害评估的建构性提升[J].灾害学,2016,31(1):188-192. [Duan H M, He Y. Constructive promotion of big data for disaster assessment[J]. Journal of Catastrophology, 2016,31(1):188-192.]
- [13] Tin P, Zin T T, Toriu T, et al. An integrated framework for disaster event analysis in big data environments[C]// Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE Computer Society, 2013:255-258.
- [14] 李学龙,龚海刚.大数据系统综述[J].中国科学:信息科学, 2015,45(1):1-44. [Li X L, Gong H G. A survey on big data systems[J]. Science China Information Sciences, 2015, 45(1):1-44.]
- [15] 程艳云,张守超,杨杨.基于大数据的时间序列异常点检测研究[J].计算机技术与发展,2016,26(5):139-144. [Cheng Y Y, Zhang S C, Yang Y. Research on time series of outlier detection based on big data[J]. Computer Technology and Development, 26(5):139-144.]
- [16] 凌骏,尹博学,李晟,等.基于监控数据的MySQL异常检测算法[J].计算机工程,2015,41(11):41-46. [Ling J, Yin B X, Li S, et al. MySQL outlier detection algorithm based on monitoring data[J]. Computer Engineering, 2015,41 (11):41-46.]
- [17] 邓敏,刘启亮,李光强.采用聚类技术探测空间异常[J].遥感学报,2010,14(5):944-958. [Deng M, Liu Q L, Li G Q. Spatial outlier detection method based on spatial clustering[J]. Journal of Remote Sensing, 2010,14(5):944-958.]
- [18] 邓敏,石岩,龚健雅,等.时空异常探测方法研究综述[J].地理与地理信息科学,2016,32(6):43-50. [Deng M, Shi Y, Gong J Y, et al. A summary of spatiotemporal outlier detection[J]. Geography and Geo- information Science, 2016,32(6):43-50.]
- [19] 葛艳琴,贾琇明.第二次土地调查建库过程中数据质量的控制方法[J].测绘科学,2008(S1):62-63. [Ge Y Q, Jia X M. Data quality control methods in the process of building databases in the second land survey[J]. Science of Surveying and Mapping, 2008,S1:62-63.]
- [20] 茅晶晶,沈红军,徐洁.全国环境统计数据审核软件设计与实现[J].环境科技,2011,24(4):65-68. [Mao J J, Shen H J, Xu J. Design and application of nationwide environmental statistics data verification software[J]. Environmental Science and Technology, 2011,24(4):65-68.]
- [21] 庄晓东,王海银,胡振彪,等.地理国情普查外业调绘核查系统实现[J].测绘科学,2016,41(2):58-61. [Zhuang X D, Wang H Y, Hu Z B, et al. Realization of iPad-based field annotation and verification system for investigation of national geographical condition[J]. Science of Surveying and Mapping, 2016,41(2):58-61.]
- [22] 薛安荣,姚林,鞠时光,等.离群点挖掘方法综述[J].计算机科学,2008,35(11):13-18. [Xue A R, Yao L, Ju S G, et al. Survey of outlier mining[J]. Computer Science, 2008,35 (11):13-18.]
- [23] 成邦文,师汉民,王齐庄.多维统计数据质量检验与异常点识别的模型与方法[J].数学的实践与认识,2003,33(4): 1-7. [Cheng B W, Shi H M, Wang Q Z. The model and method for checking quality of multidimensional statistics and identifying outliers from the data[J]. Mathematics in Practice and Theory, 2003,33(4):1-7.]
- [24] 李连发,王劲峰.地理空间数据挖掘[M].北京:科学出版社,2014. [Li L F, Wang J F. Geospatial data mining[M]. Beijing: Science Press, 2014.]
- [25] Tobler W R. A computer movie simulating urban growth in the detroit region[J]. Economic Geography, 1970,46 (Supp 1):234-240.
- [26] Anselin L. Local indicators of spatial association-LISA[J]. Geographical Analysis, 1995,27(2):93-115.
- [27] 崔鹏.中国山地灾害研究进展与未来应关注的科学问题[J].地理科学进展,2014,33(2):145-152. [Cui P. Progress and prospects in research on mountain hazards in China[J]. Progress in Geography, 2014,33(2):145-152.]
- [28] 张志彤.我国山洪灾害特点及其防治思路[J].中国水利, 2007(14):14-15. [Zhang Z T, Characteristics of mountain flood disasters in China and prevention methods[J]. China Water Resources, 2007,14:14-15.]
- [29] 郭良,刘昌军,丁留谦,等.开展全国山洪灾害调查评价的工作设想[J].中国水利,2012(23):10-12. [Guo L, Liu C J, Ding L Q, et al. Working plan for mountain flood investigation and evaluation in China[J]. China Water Resources, 2012,23:10-12.]
- [30] 黄先龙,褚明华,石劲松.我国山洪灾害调查评价工作浅析[J].中国水利,2015(9):17-18. [Huang X L, Chu M H, Shi J S. Analysis on flash flood investigation and assessment in China[J]. China Water Resources, 2015,9:17-18.]