

文章编号: 1671-8860(2004)10-0863-05

文献标识码: A

关于空间数据质量标准的若干问题

朱 庆¹ 陈松林¹ 黄 锋¹

(1 武汉大学测绘遥感信息工程国家重点实验室, 武汉市珞喻路129号, 430079)

摘要: 引入空间数据质量维的概念, 在比较分析几个标准化组织提出的空间数据质量内容组成的基础上, 结合我国的特点和现实要求, 提出了一套可供参考的空间数据质量元素和子元素。分析了空间数据质量的3种评价模式, 并将ISO/TC211的加权平均法和我国的缺陷扣分法结合, 提出了基于加权平均的缺陷扣分评价方法, 给出了这种方法的实现流程。

关键词: 空间数据; 质量标准; 质量的内容组成; 质量量度; 质量评价方法

中图法分类号: P208

由于空间现象自身存在的不稳定性、人类认识和表达能力的局限性以及空间数据在处理和使用中的变异, 使得人们对现实地理世界的抽象和表达总是存在误差。数据误差的大小直接影响着数据质量的好坏。空间数据质量关系到空间分析和操作准确度, 关系到各种空间决策支持的正确性和可靠性。缺少准确可信的数据质量评价结果, 对于数据交换、共享和使用都有很大的不利影响。当前我国地理信息数据总量很多, 而可以提供给用户使用的数据数量有限, 因而如何对空间数据质量进行全面、准确、科学的描述、度量和评价是摆在所有数据生产者和使用者面前的一个亟待解决的重要课题。目前我国已经制定了若干空间数据产品的质量评定标准, 一定程度上解决了空间数据质量评价中的部分问题, 但评价的方法和手段中不少还依赖于经验和定性的分析, 缺乏严密的理论基础和准确的定量统计分析和计算。本文就空间数据质量的内容组成和量度、评价方法两个方面讨论与空间数据质量标准相关的几个关键问题。

1 空间数据质量的内容

1.1 对空间数据质量基本概念的理解

数据质量是数据对特定用途的分析和操作的

适用程度^[1], 但这只是数据使用者的观点, 数据质量的概念对于数据生产领域和数据使用领域有着不同的含义。对于数据生产者来说, 空间数据质量是通过真实标记的原则(truth in labeling)将地理信息产品的特性和特征通过一定的方式进行标记; 对于数据使用者来说, 数据质量是按满足指定应用需求的原则(fitness for use)进行标记。例如在水文应用时, 如果格网间距为25m的数字高程模型可以满足水文应用分析的需要, 那么尽管格网间距为10m和5m的数字高程模型精度更高, 但是对于这项应用来说这三种数据的数据质量是相同的。因此, 数据质量的含义可以全面地理解为地理信息产品满足特定需要的特性和特征的总和或提供应用服务的能力^[2]。

1.2 空间数据质量的内容和空间数据质量的维

任何空间数据都存在着空间维、时间维和专题维, 因而空间数据质量也存在着这三维。空间数据包括完整性、逻辑一致性、精度和分辨率等方面的质量特性, 空间数据的质量特性和空间数据质量的维是从不同的角度观察空间数据质量的结果。它们是一种正交关系, 构成一个矩阵, 共同描述空间数据的质量。

1.3 现行标准中对空间数据质量内容的规定

国内现行的诸多标准在空间数据质量组成的命名、分类和内容的限定上都存在着较大的差异。

尽管对于不同类型的空间数据, 数据质量的具体内容不完全相同, 但由于它们统属空间数据, 因此, 它们的数据质量都存在着空间维、时间维和专题维等共性, 即不同类型的空间数据质量的内容组成都有一个共同的框架。目前正在试图通过对各种类型的数据质量内容进行归纳, 提取这一框架, 目的是希望用这个框架来演绎推理出各种不同类型的空间数据质量的内容。当然, 空间数据质量的内容体系也应具有可扩充性, 对于具体的数据类型允许将个别的差异扩充进来, 把扩充的结果作为这种数据类型的数据质量内容。尤其在当前随着 GIS 应用在众多领域的推广, 各种新的空间数据不断出现, 一旦有了这一框架, 新类型的空间数据质量的内容就比较容易得到了, 同时也保证了这些新类型的数据与已有各种数据类型的数据在评价内容上的一致性, 最终方便了数据质量的评价。

国外的一些标准化组织针对数据质量的内容做过较深入的研究, 先后提出一些数据质量元素用以标记数据质量。比较典型的有 ISO/TC211(国际标准化组织/地理信息技术委员会)、CEN/TC287(欧洲标准委员会地理信息技术委员会)和 FGDC(美国联邦地理数据委员会)等³。

虽然对于数据质量内容的描述不尽相同, 但基本都是从空间数据质量的空间维、时间维和专题维出发, 对数据质量的各个方面进行描述。

1.4 我国可以参考的空间数据质量内容

我国空间数据质量标准经过了近 20 年的发展, 在空间数据质量的具体内容上与 ISO/TC211 所制定的 ISO19113 相似, 但由于我国测绘领域的传统习惯与国外有些不同, 所采用的数据质量的元素也与其他国家有较大的差别。例如, 许多国家将定位精度分为绝对精度和相对精度, 而我国定位精度一般只考虑绝对精度。另外, 我国一些已有标准重视位置精度等定量质量元素的量化, 而在逻辑一致性、完整性、专题精度的量化上规定得不够具体, 容易导致这些方面的质量评价“凭经验”或“惯例”等, 结果降低了评价结果的可信度。

在空间数据质量三维特性的基础上, 参照国际已有和拟建标准, 考虑我国的表达习惯, 笔者提出以下质量元素和子元素用以描述空间数据质量的内容组成, 见图 1。

2 空间数据质量的量度和评价方法

2.1 空间数据质量的量度

空间数据质量的量度是测试和记录空间数据

空间数据质量元素		空间数据质量子元素	
定量 质量 信息	完备性	空间维	地理范围的完备性
		数据分层的完备性	
		实体类型完备性	
		专题维	属性数据完备性
			注记完备性
	逻辑一致性		要素间关系完备性
		时间维	时间维时间的有效性(现势性)
		空间维	拓扑关系的一致性
	精度	专题维	格式的一致性
		时间维	数据采集和生产时间的一致性
		空间维	数学基础
	分辨率	平面位置精度	
		高程精度	
		接边精度	
	分辨率	专题维	属性的正确性
			注记的正确性
		时间维	时间精度
非定量 质量 信息	空间维	空间分辨率	
	专题维	分类与代码的正确性	
	目的 用途 数据志	数据分层的合理性	
		时间维	时间分辨率

图 1 可供参考使用的空间数据质量的内容组成

Fig. 1 Geospatial Data Quality Components for Reference

质量结果的方式, 常用的量度有以下四种^[4]。

- 1) 合格/不合格;
- 2) 错误的绝对数量;
- 3) 错误的相对数量(错误的数据项个数/数据项的总数);
- 4) 统计描述(如精度值的描述等)。

一定的数据质量元素和子元素对应一定的数据质量量度。空间数据质量评价时, 可根据实际用途的不同选用以上 4 种量度之一。例如, 对于定位精度的检测评价, 其精度值需要定量的记录, 质量量度可选用量度 4); 对于逻辑一致性的检测评价, 错误的数量需要统计计算, 质量量度的方式可选用量度 2) 或量度 3); 对于时间精度的评价则通常只需要判断其是否合格, 因此可选用量度 1)。

2.2 空间数据质量评价的 3 种模式

目前的空间数据质量评价存在 3 种模式, 即最小质量标准模式、元数据标准模式和市场标准模式^[5]。

- 1) 最小质量标准模式。它是一种规定阈值的评价模式。在质量评价由数据生产者来完成的情况下, 这是一种质量控制形式。它基于对数据的检测来确定是否符合预先定义的阈值要求来评价数据的质量。

2) 元数据标准模式。它是一种真实标记模式。这种模式把错误看成是不可避免的, 不是在质量评价时强加一个最小的质量标准, 而是由数据的使用者来评价数据的适用性, 数据生产者的责任是编写元数据文档, 真实标记错误或误差的情况。这种模式比较灵活, 但是数据的使用者不能反馈相关的质量信息, 也就是说这还是一种单向的信息流, 它制约着数据生产者纠正数据中的错误。

3) 市场标准模式。它是一种真实标记和信息反馈结合的模式。这种模式利用双向的信息流来获取数据使用者对数据质量问题的反馈, 一方面, 数据生产者将数据质量评价的结果真实标记于元数据中, 随数据一起发布给数据使用者; 另一方面, 数据使用者也将数据使用过程中发现的数据中存在的质量问题反馈给数据生产者。数据使用者的反馈信息将对数据中存在的质量问题的修复起到重要作用。一个典型的例子就是微软的反馈向导的使用, 这种软件便于用户通过电子邮件反馈发布的空间数据中可能存在的质量问题。这种模式在市场环境下很有用, 它能提高数据生产对用户的要求和期望的满足程度。

2.3 空间数据质量的评价方法

2.3.1 评价方法的分类

空间数据质量的评价方法可以分成直接评价方法和间接评价方法^[6]。直接评价方法是通过对数据集抽样并将抽样数据与各项参考信息(评价指标)进行比较, 最后统计得出数据质量结果; 间接评价方法则是根据数据源的质量和数据的处理过程推断其数据质量结果, 其中要用到各种误差传播数学模型。杜道生等针对两种评价方法的特点提出了直接评价和间接评价的质量模型^[2]。

间接评价方法是从已知的数据质量计算推断未知的数据质量水平, 某些情况下还可避免直接评价中繁琐的数据抽样工作, 效率较高。针对数据质量的间接评价, 不少学者基于概率论、模糊数学、证据数学理论和空间统计理论等提出了一些误差传播数学模型, 但这些模型的应用必须满足一些适用条件, 总的来说, 要想广泛准确应用这些误差传播的数据模型来计算数据质量的结果, 目前还存在较大难度, 因此, 间接的评价方法目前应用还较少。在数据质量的评价实践中, 国内应用较多的是直接评价方法。

2.3.2 直接评价方法中的缺陷扣分法

在我国已有的一些数据质量标准中, 直接评价方法大多使用缺陷扣分法, 这种方法既有利也

有弊。我国的缺陷扣分法包含了对数据中的不同类型错误或误差按照其对数据质量影响程度的大小进行加权统计的思想, 虽然在缺陷级别个数的设置和各缺陷级别的扣分量的设置上存在较大的灵活性, 但应用范围较广, 适用于各种类型的数据质量评价。该方法操作的难度主要在于数据质量评价指标的制定上, 一方面, 错误类型繁多, 多个相同类型的错误一起才可以记作一个某级别缺陷, 这需要多次实验才能得出一个合适的值; 另一方面, 每种缺陷对应的扣分数量取多少合适, 也需要多次实验才能得出结论。

2.3.3 ISO/TC211 加权平均法

ISO/TC211 的加权平均法也属于数据质量直接评价的方法。该方法的实现过程是首先选择适用的数据质量元素和子元素, 并将数据集按照特征分成若干地物要素(如居民地、道路、水系、植被等), 给每一种地物要素按照其在数据集中的重要性分配一个适当的权重(大小为 0.0~1.0, 权重总和为 1.0), 然后给每个数据质量元素选择一种数据质量量度, 再对数据集中的每一种地物要素进行抽样, 统计该地物要素中错误数据的总量占抽样数据的百分率, 得出数据集各地物要素的正确率 C_i , 最后按照各地物要素的权重计算其加权平均, 并把它作为数据质量的结果值:

$$R = \sum_{i=1}^k (C_i \cdot W_i) \quad (1)$$

式中, R 为数据质量结果值, $R \in (0.0 \sim 1.0)$; C_i 为第 i 种地物要素数据的正确率, $C_i \in (0.0 \sim 1.0)$; W_i 为第 i 种地物要素对应的权重, $W_i \in (0.0 \sim 1.0)$; k 为地物要素类型的总数。

2.3.4 基于加权平均的缺陷扣分评价方法

基于加权平均的缺陷扣分评价方法是对我国的缺陷扣分法和 ISO/TC211 的加权平均法的融合, 它既考虑同一种地物要素中不同缺陷级别的错误对数据质量结果所产生的影响程度不同, 也考虑由于不同地物要素本身在整个数据集中的重要程度不同, 而造成这些地物要素中的错误对数据集质量的影响程度不同, 因此评价的结果较上面的两种方法更准确, 但操作过程比上面两种方法复杂。在给出基于数据质量得分的数据质量分级方案的情况下, 它也可以评价出数据集的质量等级。数据集各要素的缺陷的个数、级别、数据质量结果值和数据质量等级共同构成了数据质量评价报告。其实现过程如图 2 所示。

该方法数据质量结果值的计算公式为:

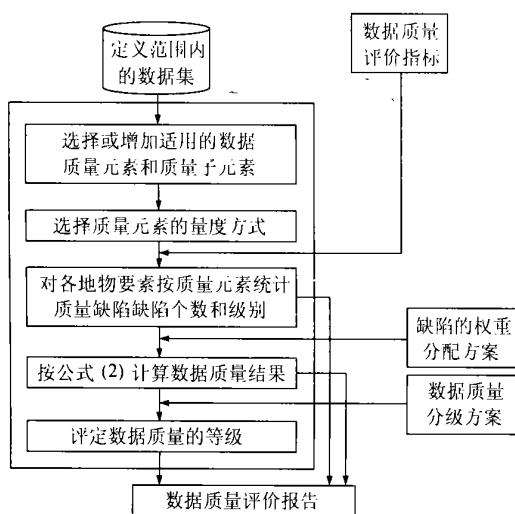


图 2 数据质量等级评定流程图

Fig. 2 Flow chart of Evaluating Data Quality Levels

$$N = T - \sum_{j=1}^f [W_j \sum_{i=1}^k (n_i \cdot s_i)] \quad (2)$$

式中, N 为数据质量的得分; T 为数据集预置得分; W_j 为第 j 种地物要素对应的权重; f 为地物要素总数; n_i 为缺陷级别为 i 的缺陷个数; s_i 为缺陷级别为 i 的一个缺陷对应的扣分; k 为缺陷分级的数目。

3 结语

本文在对空间数据质量内容的描述中, 得出了我国适宜采用的空间数据质量内容(包括数据质量元素和子元素)。融合我国普遍采用的缺陷扣分法和 ISO/TC211 的加权平均法, 可以进一步明确空间数据质量的含义、内容、分类、分级、质量的评价指标和评价方法等, 适用于基础地理信息和各种专题信息的质量控制、检查验收与质量评定, 具有较广泛的适用性和较强的可操作性。

随着空间数据质量相关研究的不断深入, 空

间数据质量的量度和评价的方法日趋科学化。由于空间数据质量的评价工作量很大, 实现空间数据质量评价的自动化有着重要的意义。通过开发软件来代替繁琐的目视判读, 依靠特定算法实现对数据中存在错误或者误差进行量测、统计和计算已成为可能, 这不仅可以提高数据质量评价的效率, 还可以减小数据质量评价中人为因素造成的误差, 提高数据质量评价的质量。

致谢: 感谢杜道生教授在本文写作过程中给予的指导和支持。

参 考 文 献

- Howard V. Data Quality Measurement and Assessment. NCGIA Core Curriculum in GIScience. <http://www.ncgia.ucsb.edu/education/curriculumgiscc/units/u100/u100-f.html>, 1998
- 杜道生. 空间数据质量模型研究. 中国图像图形学报, 2000, 5(7): 559~562
- 童小华. GIS 数据误差处理与质量控制理论与方法: [博士后出站研究报告]. 武汉: 武汉大学, 2003
- ISO 19113. Geographic Information-Quality Principles. <http://www.statkart.no/isotc211/>, 2001
- Gerhard J. Standardization of Data Quality Measures. Proceedings of the 2nd International Symposium on Spatial Data Quality' 03 Hong Kong, 2003
- 国家测绘局. 数字测绘产品检查验收规定和质量评定. 北京: 测绘出版社, 2001
- 阎正, 蒋景乾, 杜道生, 等. 城市地理信息系统标准化指南. 北京: 科学出版社, 1998
- DIS 19114. Geographic Information-Quality Procedures. <http://www.statkart.no/isotc211/>, 2002

第一作者简介: 朱庆, 教授, 博士生导师, 现主要从事数字摄影测量、虚拟地理环境和数码城市 GIS 的理论、技术方法和应用研究。代表成果:《数字高程模型》和数码城市 GIS 软件 CCGIS。E-mail: zhuq66@263.net

Key Issues on Quality Standardization of Geospatial Data

ZHU Qing¹ CHEN Songlin¹ HUANG Duo¹

(1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: On the basis of the concept of geospatial data quality, this paper introduces the dimensions of geospatial data quality for further understanding the components. The quality dimensions include spatial dimension, temporal dimension and thematic dimension. For each dimension, there are several quality components. A series of geospatial data non-quantitative quality elements, quantitative quality elements and subelements are proposed as a reference of geospatial data

quality components standardization for China. This paper discusses measures and evaluation methods for geospatial data quality and lists four types of geospatial data quality measures and gives a simple guide about how to select a suitable measure for a subelement. Through combining the weighted and aggregated quality evaluation method of ISO/TC211 and negative scores quality evaluation method based on quality effect in China, a kind of new quality evaluation method is introduced, and the operation procedures is illustrated.

Key words: geospatial data; quality standard; quality components; quality measures; quality evaluation method

About the first author: ZHU Qing, professor, Ph. D supervisor. His research interest is in the theoretical and technical research of digital photogrammetry, virtual geographic environment and cybercity GIS. His representative achievements include a book titled Digital Elevation Model and CCGIS software.

E-mail: zhuq66@263.net

(责任编辑: 涓涓)

武汉大学参加北极黄河站首次科学考察

武汉大学派出的中国南极测绘研究中心鄂栋臣教授和青年教师艾松涛,在参加中国北极黄河站首次北极科学考察队,圆满完成考察任务后,已于8月25日返汉。

中国北极黄河站首次科学考察队,经过国家海洋局极地办组织专家论证、遴选,由包含测绘遥感、地质、环境、冰川、海洋、大气等6个学科的专家组成。经国务院批准,首次科考队包括鄂栋臣和艾松涛在内的11名队员,于2004年7月20日从北京出发。中国北极黄河站,位于挪威北部斯瓦尔巴群岛的新奥尔松北极站区,地理坐标为北纬 $78^{\circ}55'$,东经 $11^{\circ}56'$ 。2003年我国与新奥尔松当地的挪威王湾公司签约,筹备建立中国北极科考站,2004年7月28日正式揭牌投入运行。

自从1984年我国进行首次南极考察以来,武汉大学中国南极测绘研究中心就承担着我国南北两极的测绘科考重点项目。中心主任鄂栋臣教授曾经5次奔赴南极,三次到达北极进行科考。近20年来,鄂栋臣教授一直主持我国的南北两极测绘保障研究项目,积累了丰富的极地考察经验和成果,取得多项省部级、国家级奖励。此次北极黄河站的科学考察,是鄂栋臣教授主持的北极卫星跟踪站和动态信息系统建设项目,是充分利用我国北极建站和极地科学考察向北极地区扩展的有利条件,建成我国北极地区GPS卫星常年跟踪站,结合南极长城站、中山站GPS卫星跟踪观测,与我国境内的卫星观测站构成可跟踪观测长弧段的观测网,扩展了我国卫星跟踪轨道范围,提高了两极卫星观测数据在空间技术中的应用研究效益,并为参与国际合作研究创造了良好条件。

目前,北极黄河站所在的新奥尔松已经有8个国家建立了11个科考站。我国在北极地区建站晚,科考历史短,因此必须在了解本学科领域当前国际北极研究水平的基础上去寻找研究的新起点,尤其是极地地球动力学等现代极地测地学问题研究,周期长,规模大,范围广,必须多渠道去寻求国际交流与合作,以实现数据、信息等共享,使武汉大学在国际北极科考研究中能获得一席之地。