

文章编号:1001-9081(2011)10-2849-04

doi:10.3724/SP.J.1087.2011.02849

GIS 矢量数据制图规范性质量评价系统构建与应用

曹俊^{1,2},廖顺宝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. 中国科学院研究生院, 北京 100049)

(caoj@lreis.ac.cn)

摘要:以全国1:25万土地利用/覆被数据中存在的规范性质量问题为研究对象,利用开源GIS软件构建一个操作简易、评判客观、全面的规范性质量评价系统。这种质量检查系统是基于规则的新型质量评价体系,针对不同的数据质量问题灵活地设计出质量检查规则,并有效集成到系统当中。实验结果表明,系统能灵活、高效地检查出土地利用/覆被数据中的规范性质量问题,并给出客观的评价。

关键词:地理信息系统;数据质量;规范化;评价模型;指标库

中图分类号: TP311.51 文献标志码:A

Construction and application of quality evaluation system of GIS vector data normative mapping

CAO Jun^{1,2}, LIAO Shun-bao¹

(1. State Key Laboratory of Resources and Environmental Information System,
Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Regarding the existing normative data quality issues of the national land use/cover at scale of 1 to 250 000 as the research object, the open source Geographical Information System (GIS) software was used to build a normative quality evaluation system that had characteristics of simple operation, objective evaluation and comprehensive inspection. This quality inspection system is a new rules-based quality evaluation system. According to different data quality issues, it flexibly designed quality inspection rules, and effectively integrated into the system. The experimental results show that the system can flexibly and efficiently check out the land use/cover normative data quality issues, and give an objective evaluation.

Key words: Geographical Information System (GIS); data quality; normalization; evaluation model; index base

0 引言

随着地图制图技术的发展,从手工数字化方式逐步转变为遥感与专家系统辅助制图,效率越来越高,制图的范围也是以往制图方式无法比拟的。20世纪80年代国家1:100万土地覆被地图集编制花费了近十年时间;相比而言,2005年全国土地覆被数据制图工作为期一年,生产效率提高10倍^[1]。中国科学院遥感研究所采用的是遥感影像自动分类结合矢量图进行地类识别^[2],在保证一定精确度的基础上,缩短生产周期,但是在产品质量检查过程中也发现了大量的由制图系统本身的不完善而带来的错误。这些制图错误的产生来源于制图综合技术的不完备,并且一般肉眼难以全部发现。在全国1:25万土地利用/覆被质量检查任务当中发现了典型质量问题有如下几种:土地利用图斑属性值错误、碎屑多边形过多、相邻图符信息承载量相差悬殊、过度综合后规则多边形太多等,综合来说称之为规范性质量问题^[3],制图不符合规范。

一些学者和生产单位根据生产任务的需求自主编制了相关的质检系统,具有对一般质量问题的检查功能,而对于规范性质量问题的检查与评价,目前尚无专用的检查系统。这些系统中,检查规则固化在程序里,难以灵活扩展,而且是基于GIS平台的二次开发,系统体积较大,操作复杂。本文在综合

已有质量检查软件的功能特性的基础上构建出具有以下优势的规范性质量检查系统:1) 使用体积较小的开源GIS软件作为基础,降低系统依赖性;2) 开发地图操作中间件,结合脚本引擎技术,将质检算法写入脚本中,方便使用者灵活添加质检规则;3) 针对规范性质量问题设计检查算法。制图规范性质量检查需要根据质量问题的特性设计检查算法,下面简要介绍典型的制图规范性质量问题的表现。

1 土地覆被数据常见的问题

1.1 土地利用/覆被数据属性值错误

属性值错误主要是指图斑所代表的地物类型的编码不正确。表现为两种情况:第一是错类,由遥感制图时的分类算法产生的错误类别赋值,根据全国土地覆被相关制图标准,1:25万图上农田、聚落定性判读精度要求为90%,其他80%以上;另一种是通过属性编辑后产生的错类,编码值不属于常规的地类编码,是根本不存在的属性值。

1.2 碎屑多边形不符合制图规范

根据制图标准,不同的地类图斑在对应的制图比例尺下应该存在最小面积阈值,《中国1:25万土地覆盖遥感制图技术实施方案》中规定1:25万分幅或标准分幅图上最小图斑,农田、聚落4mm²,其他类型9mm²;狭长图斑最短边下限2mm^[2]。低于面积标准的图斑对数据整体影响不大,可以合

收稿日期:2011-04-15;修回日期:2011-06-13。

作者简介:曹俊(1986-),男,湖北荆州人,硕士研究生,主要研究方向:GIS矢量数据质量评价;廖顺宝(1966-),男,四川德阳人,副研究员,博士,主要研究方向:遥感与地理信息系统、地学数据的开发与集成、系统设计与开发。

并到相邻图斑上,或者视为无效,应予以删除。

表 1 全国土地覆盖遥感监测主要制图指标

内容	要求
最小图斑	1: 25 万分省或标准分幅图上农田、聚落 4 mm^2 , 其他类型 9 mm^2
狭长图斑最短边宽	相同图上 2 mm
图斑定性判读精度	1: 25 万图上农田、聚落 90%, 其他 80% 以上
图斑定位转绘精度	人机交互判读精度的一般偏差 $< 0.6 \text{ mm}$, 最大 1.0 mm

1.3 相邻图符信息承载量相差悬殊

相邻图符接边处为两个图幅的过渡带,两者之间的信息量应该处于相当的水平,如果出现图斑出现稀疏/浓密对比差异,即能引起数据信息负载量的差异的质疑,进而对数据质量的可靠性产生疑问。该类问题是相邻图幅之间的数据源、制图流程差异性带来的。

1.4 过度综合后规则多边形

计算机制图在提高效率的同时,简化了人为操作的判断,控制人为常犯的错误后,忽略了其他类型的错误。地图综合还没有发展到智能化的阶段,结果的可靠性经常不够,表现为地图上出现某一个比例尺下面超出一定长度的近似于直线、直角的形状,然而自然地物形状出现大范围的规则图形的可能性较低。

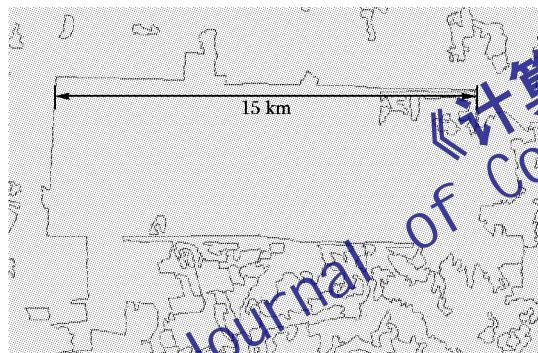


图 1 制图综合中出现的较规则的形状

上述 4 类错误是实际生产中常见的错误,这些错误主要是制图和分类过程中没有严格遵循制图规范造成的,不涉及图形的位置精度问题,一并称之为矢量数据规范性质量问题^[4]。人们在追求用计算机代替人进行地图制图的同时,也在积极思考利用电脑代替人进行质量检查的方法。如何对这些数据生产过程中产生的错误进行分析归纳,设计解决方案,是本文本要探讨的主要内容;建立相应的质量检查模型、设定质检规则,并且使其在同一系统里运行,最终得到检查结果,根据检查结果进行评分^[3,5-7]。

2 系统体系结构与功能模块设计

2.1 质量检查方式

目前矢量数据质量检查的基本方法包括屏幕显示对比、绘图叠合检查、利用 GIS 辅助软件检查、打印校对、手工评价等必须人工参与的检验方法^[8]。随着制图方式逐渐变得自动化,生产效率越来越高,相应地,也要提高质量评价的效率。如何利用计算机自动化检验,成为了矢量数据质量控制与评价的研究重点,值得深入探讨。

2.2 自动化检查系统的建立流程

建立自动化数据质量检查评价系统的工作流程是:可行

性分析、系统设计、详细设计、数据验证、成果评价,比一般软件工程的工作流程要多一步,对数据实例进行检查标准定制化。不同的专题数据,使用的标准是不一样的,但是系统可以根据数据进行检查规则的自定义,达到在一个系统中使用不同的检查标准。

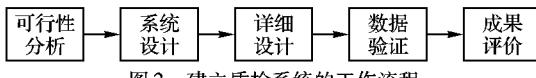


图 2 建立质检系统的工作流程

检查标准是根据所要检查的数据产品来搜集和整理的,比如 1: 25 万土地覆被数据,就需要用到 1: 25 万土地覆被数据生产规范、分类标准、检查技术指标等规范文件。根据人工抽样检查出来的质量问题,进行归类分析,并抽象出检查的规则,编制对应的检查程序,即所谓的检查规则^[5]。

质量检查评价系统的功能组成:该系统具备数据的读取和显示功能、图符的浏览漫游、数据矢量元素的遍历访问与检查、检查结果的记录、问题几何元素的定位与检查结果查看、结果报表生成、质量评价等。

2.3 系统分层搭建与模块分解

分层设计质量检查系统,如图 3 所示。首先是数据访问层,中间为业务层(即运行规则进行质检判断的一层),最上层为结果展现层。

数据访问层包括抽象数据访问接口及各种数据源提供对象组成。该层的主要功能是对于支持的数据格式,在对应的数据提供者对象的访问时,为访问接口提供统一的中间格式,使用的是开放地理信息系统协会(Open GIS Consortium, OGC)所支持的标准格式。业务逻辑层包括质量检查规则库以及质检结果库,将质检规则编写成脚本程序,嵌入到质检系统当中,质检时根据数据特有的质量问题进行配置,做到针对不同的专题数据有不同的检查方案。质量检验结果存入数据库中,按不同的问题逐条记录错误,方便问题的检索与核对;结果展现层结合 GIS 数据显示功能,将不同的质量问题显示到屏幕上,使检查者有直观的认识,方便对质量问题查找原因和修改。

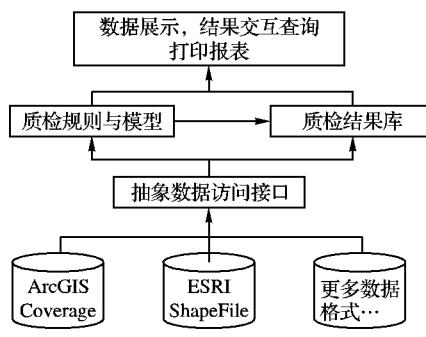


图 3 质检系统的三层体系结构

2.4 数据层设计及 JMap 组件开发

数据访问层往下细分为数据驱动层和高级编程接口层。数据驱动使用了 GDAL/OGR 库中矢量数据读写程序,同时将此软件的数据读写接口封装成组件对象模型(Component Object Model, COM)接口,提供给 VB 脚本语言访问。GDAL/OGR 库目前由 OSGeo 开放社区(<http://www.gdal.org>)维护,开放源代码,支持常用的栅格和矢量格式数十种,包括 ESRI Shapefile、ESRI ArcSDE、GML 等 53 种矢量数据格式以及 Arc/Info ASCII Grid、GeoTiff 等 117 种栅格数据。跨平台、开放性、支持多语言等好处让科研人员爱不释手。在 GDAL/OGR 数

据访问层上进行封装和抽象,建立具有工业标准的开放接口,可以支持 VB、C++、CSharp 等语言无关的 COM 控件 JMap ActiveX V1.0(以下简称 JMap 组件),该组件作为数据操作的中间件,既能使用 GDAL/OGR 强大的数据读写功能,又能使用脚本语言,达到系统可扩展的要求。

OCR 几何对象 C++ 模型是按 OGC 简单对象模型(如图 4 所示)编写的,是对 OGC 模型的一个实现版本。选择 OGR 矢量数据操作程序库作为数据驱动程序极大地缩短了质量检查系统的开发时间,并且能保证数据访问层面的正确性。OCR 提供的 C++ 对象派生关系如图 5 所示。

JMap 组件的开发是整个系统开发的核心工作之一。该控件是连接数据读取、显示、质检功能的纽带,贯穿了三层结构的

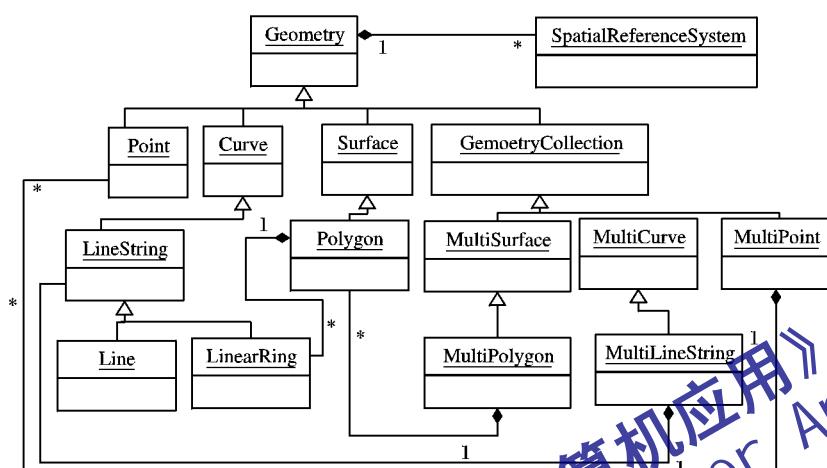


图 4 OGC 矢量几何对象 UML 模型

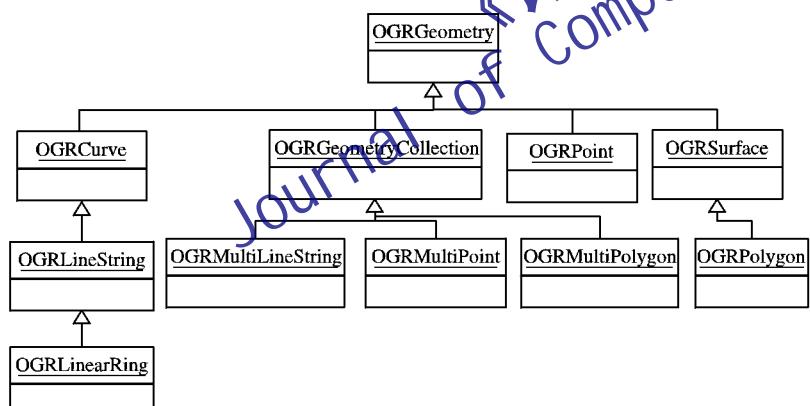


图 5 OGR 几何对象的 C++ 对象派生关系

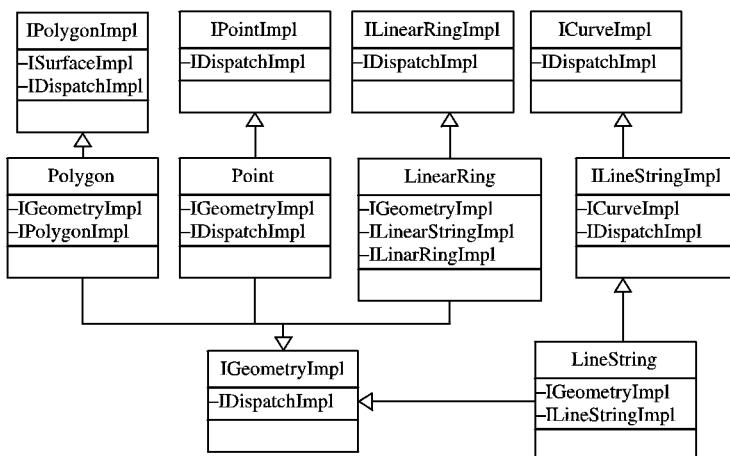


图 6 JMap 组件中的几何数据操作对象模型

始终;控件的开发借鉴了 ArcGIS 可自定义脚本的软件设计模式,利用 COM 技术将底层的数据驱动和数据展示功能包装到一个组件库中,在系统中集成 VB 脚本引擎,为系统提供二次开发功能。JMap 组件基本功能包括地图数据逻辑组织、绘图、浏览等操作。JMap 组件几何图形对象模型如图 6 所示。

JMap 采用 COM 组件对象模型编程模式,封装了从数据操作至显示在内的基本功能,是一套 COM 对象集合,包括数据源提供对象 Provider、几何对象读取接口 IGeometry、图层对象 Layer、地图对象 Map 等。底层的数据读取操作来自于 ShapeFile 读写开源程序库,IProvider 接口提供以 IGeometry 为接口的对象读取函数,为 Layer 提供数据访问方式,同时 VectorRenderer 矢量渲染类可以根据 IGeometry 接口返回的几

何对象结合 Style 样式渲染图片,由 Map 对象返回给 MapControl 对象进行屏幕显示。JMap 组件对象自下而上的逻辑关系模型如图 7 所示。

2.5 质检数据库模型设计

质检数据库模型是指在系统中如何组织数据,如何为同一类数据制定检验规则的一套数据库表格。例如 2005 年全国土地覆被数据有 32 个省、市、自治区的 Coverage 数据集(每个图集中包含点、线、面图层),数据量大,内容丰富。如何去管理这些数据,成了数据组织的难题。对于这类数据(32 个数据集)可以定义一个类别,称之为标准(Standard),即“2005 年全国土地覆被数据集”;一个数据集下有 3 个图层,每个图层下面又有字段定义,构成“标准→图层→字段”的表空间,共同描述要考查的数据对象。

设计了 DicStandard、DicStandardType、DicLayer 和 DicField 共 4 个表格作为管理数据的表空间。DicStandard 表是为某一类专题数据定制字典名称,主键是 StandardID,表示该类数据的唯一标识符;DicStandardType 表是对 DicStandard 表的描述,说明标准的名称,描述等信息;DicLayer 表记录了所有图层信息,主键是 LayerID,外键是 StandardID;DicField 表记录了所有字段信息。4 个表之间的关联关系如图 8 所示。

系统数据库同时记录下质检任务的配置情况。配置质检任务是在数据“标准→图层→字段”表结构的基础上,构建出基于“质检任务→数据集→规则集”的对应关系。质检任务是对一组相关数据的整合,包含了待检数据集,一对多关系;每个数据集包含多个图层;每个图层关联多个质检规则。3 个一对多关系共同构成了质检任务配置。

2.6 质检算法的建立与运行机制

为方便质检人员灵活制定检查方案,质检系统嵌入脚本引擎,采用二次可编程的方式将质量检查规则编写成 VB Script 脚本,集成到规则库当中。检查规则由质检算法组成,是根据同一类型的数据,查阅相关的数据

质量检查规范、生产指标,编写出对应的检查项程序模块,每一个质检算法都是依附于质检系统的 VB 代码,包括取出数据、进行指标项检查,以及把结果写入数据库等一系列工作。VB 脚本在 VB 脚本引擎下执行,用户的主要工作是通过配置界面为数据配置检查项,然后统一地执行检查项,将质检结果记录到数据库中。

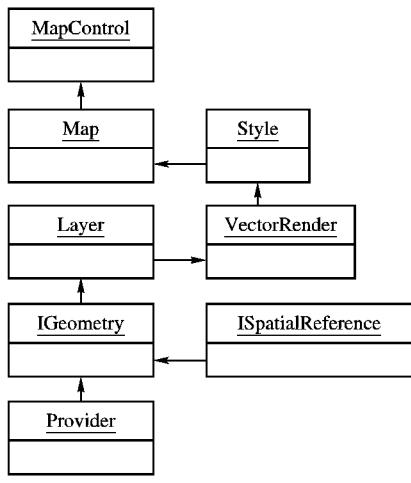


图 7 JMap 组件中的逻辑结构模型

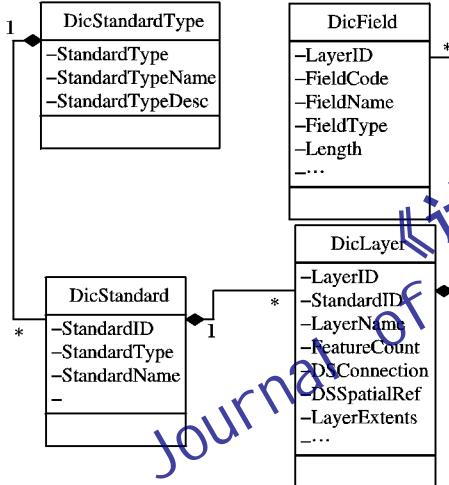


图 8 质检数据库模型设计 UML

2.7 地图浏览与结果交互查询

考虑到系统的易用性,对用户界面进行了简化处理,采用微软基础类库(Microsoft Foundation Classes, MFC)开发出可视化框架,将 MapControl 控件嵌入在 MFC 框架的视图当中,能够显示由 JMap 组件类库中渲染出来的图片,同时在选择放大、缩小、平移等操作模式时,会是地图响应的操作。系统读取数据并显示的效果如图 9 所示。



图 9 质量检查系统界面

质量检查任务运行完成后,可以查看每个数据集对应的

质量检查结果,出现的问题与相应的图形关联在一起,存储到数据表里。在用户界面中,可以通过选择问题列表的一行记录,定位到该图形上,帮助核检者直观的判断问题出现的原因并且快速修改。图形质量问题检核界面如图 10 所示。

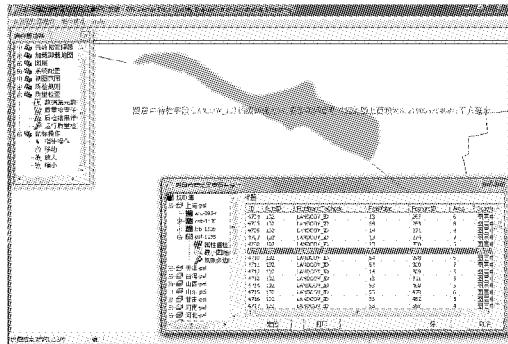


图 10 图形问题检核界面

3 结语

本文讨论了利用开源 GIS 软件来构建矢量数据规范性质量检查系统,通过分层搭建质检系统来降低系统的复杂度。

系统的重点在于 JMap 组件的研发工作,该 COM 控件集 GIS 数据源读写、几何对象控制、界面展示等功能于一体,贯穿了三层结构;另外基于 VB 脚本引擎的质检规则的编写,将质检规则单独的写成 VB 脚本,可以方便地为不同的数据进行定制,达到一次编写、重复利用的功能,提高了开发效率;最后是数据的组织功能,通过质检数据库体系,将待检数据导入到数据表中,通过关系表将质检规则和数据集关联起来,执行质检任务之后,再用关系表将质检结果和数据、规则关联起来,并通过统一的用户界面进行错误的查看和定位。本文主要关注如何将质检系统搭建起来,而如何根据数据不同的错误和检查规范来编写质检算法,没有做详细的探讨,这是以后工作的重点。

通过该系统的研制,能够在一定程度上提高数据质量检查的可控性,提高效率,特别是在数据中心检查数据时,将已有的质检规范转化成质检规则,固化成程序模块,直接应用到数据集本身,为数据制定一份质检报告,并作为数据集元数据的一部分,这样对数据的更新非常有利,同时也为数据使用者提供了更多的质量参考信息,进而为产出高可靠性的科研成果提供了有力保障。

参考文献:

- [1] 宋乃平,张凤荣,王磊,等.我国土地利用/覆被研究的热点与应用发展问题[J].地球信息科学,2008,10(1):60~66.
- [2] 张增祥,汪潇,王长耀,等.基于框架数据控制的全国土地覆盖遥感制图研究[J].地球信息科学学报,2009,11(2):216~224.
- [3] 廖顺宝,曹俊.数字地图规范性质量评价探讨[J].浙江大学学报:理学版,2010,37(5):583~587.
- [4] 薛芳.数字线划图(DLG)的质量评定方法[J].城市勘测,2003(1):42~46.
- [5] 张彦彦.基于规则的 DLG 数据质量检查方法研究[D].南京:南京师范大学,2007.
- [6] 王帆飞.空间数据库数据质量评测与质量控制体系研究[D].成都:四川大学,2005.
- [7] 车兴锋.4D Checker 数字测绘产品质量检验系统的研制与开发[J].测绘通报,2004(9):47~65.
- [8] 曾衍伟.空间数据质量控制与评价技术体系研究[D].武汉:武汉大学,2004.