

顾及地理要素变化过程的数据增量更新方法*

罗国玮^{1,2}, 张新长¹, 齐立新¹

(1. 中山大学地理科学与规划学院//广东省城市化与地理环境空间模拟重点实验室, 广东 广州 510275;
2. 广西师范学院, 广西 南宁 530001)

摘要:为解决空间矢量数据更新前后数据之间的关联关系识别困难的问题,在基于要素更新的基础上,引入了变化过程库,提出了一种顾及要素变化过程的更新方法,对由更新引起的要素变化信息进行识别并存储,设计了变化信息库的数据结构并对要素变化信息的应用作了说明,并以某市地形图数据在数据更新系统中做了试验来验证所提出的方法。试验结果表明,改进后的更新方法更利于历史数据的回溯及地物生命周期的跟踪。

关键词:变化过程;增量更新;缓冲区;历史数据

中图分类号: P208 **文献标志码:** A **文章编号:** 0529-6579(2014)04-0131-06

A Incremental Updating Method of Spatial Data Considering the Geographic Features Change Process

LUO Guowei^{1,2}, ZHANG Xinchang¹, QI Lixin¹

(1. School of Geography and Planning//Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou 510275, China;
2. Guangxi Teachers Education University, Nanning 530001, China)

Abstract: The expression of link information between updating data and history data is still a difficult issue so far. This paper proposes a way to identify and store the feature changing information caused by data updating process based on the consideration of features'changing process. Besides, this paper designs the record structure of changing information and verifies the method by real data in a prototype system. The experiment results show that the improved method has better performance in the process of geometric history data backdating and recording of geographic object life cycle.

Key words: change process; incremental update; buffer; historical data

数据更新是保持空间数据现势性的主要方法,为节省数据采集和更新的时间,目前大多矢量数据更新采用增量更新方法,只更新发生变化了的数据。文献[1]提出了基于拓扑联动的更新方法,该方法比较适合地籍数据库的增量更新。文献[2]提出了一种自适应的矢量数据增量更新方法,通过范围内要素的替换,并进行接边操作,实现增量更新;文献[3]提出基态修正的GIS数据库增量更新方法,设计了增量信息的存储模型,但这些

方法没有对地理要素的变化信息进行记录存储。另外还有基于事件及面向时空过程的更新方法^[4-6],对触发地物发生变化的事件进行了记录存储,但这类方法面向业务主体构建,且对地理事件信息的完整性要求高。目前对矢量空间数据库增量更新技术的研究包括增量信息的存储模型、增量信息发现和历史数据的组织问题,以及版本生成与时空信息检索等^[2],但很少有对变化信息的表达及存储模型进行研究。

* 收稿日期: 2013-09-27

基金项目: 国家863计划基金资助项目(2013AA122302); 高等学校博士点专项基金资助项目(20120171110030)

作者简介: 罗国玮(1979年生),男;研究方向:空间数据更新与融合;通讯作者:张新长;E-mail: eeszxc@mail.sysu.edu.cn

数据更新操作改变了空间数据库的内容, 增量更新只对变化了的的空间信息进行操作^[7], 更新后的地理要素与更新前相比会发生变化, 这些要素的变化过程对了解地物的生命周期及发展趋势有重要的意义^[8]。目前的矢量数据增量更新方法大多只是将被替换数据存入历史库, 并没有记录由于更新导致的地理要素变化信息, 多次更新导致要素发生变化后, 很难识别新旧数据之间的要素关联关系。本文提出了在矢量数据增量更新中加入变化过程库对地理要素的变化过程进行存储, 设计了通用的地理要素的变化信息存储模型, 并对地理要素变化过程库的应用做了示例说明。

1 增量更新方法

1.1 更新流程

本文研究的增量更新方法是针对同级比例尺的矢量空间数据。目前, 常用的空间数据增量更新有基于范围的更新方法和基于要素的更新方法^[9]。基于范围的更新方法是将旧数据按相应的范围挖空后将新数据填入, 然后进行要素接边操作。基于要素的更新是对新旧数据的要素进行变化检测, 只更新变化了的要素。为便于在数据更新的同时能记录地物的变化过程, 本文采用基于要素的更新方法。更新流程如图 1 所示。更新工作主要在临时库中进行, 首先要对新旧数据做要素匹配及变化检测, 当检测到变化信息后, 再对旧数据进行添加、删除及修改操作, 进行相应的冲突检测后, 将要素的变化信息存入变化过程库, 被修改和删除的旧要素存入历史库。

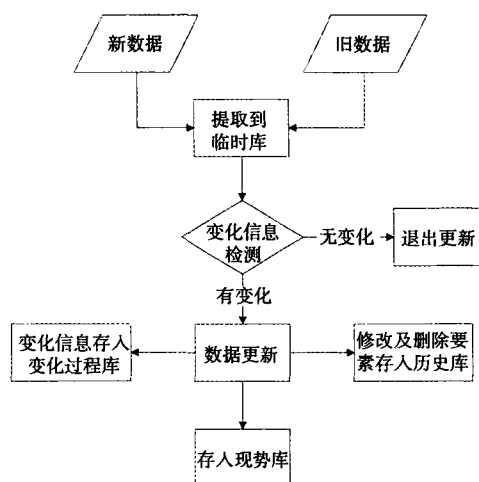


图 1 矢量数据更新流程

Fig. 1 The process vector data update

1.2 变化信息检测

变化信息检测的目的是为了发现增量信息, 将新数据与旧数据进行要素匹配。对于不同类型要素的变化检测采用不同的方法进行, 点状要素采用欧式距离及属性信息^[10]进行匹配, 线状要素的匹配与变化检测^[11]从几何上采用缓冲区分析、交叉面、hausdorff 距离^[12]、概率统计^[13-14]等方法结合属性信息进行, 面状要素的匹配与变化检测可以通过重心距离、重叠面积结合属性信息进行。本文以面要素为例, 采用空间关联度结合属性语义关联度进行要素变化检测。设 A_i ($i=0, 1, 2 \dots m$) 为新要素, B_j ($j=0, 1, 2 \dots n$) 为旧要素。先以新要素 A_i 生成缓冲区, 获取与 A_i 缓冲区相交的候选要素 B_j , 空间关联度 $P(A_i, B_j)$ 通过计算 A_i, B_j 要素缓冲区的交集与并集的面积比得到, 计算方法如公式 (1) 所示。

$$P(A_i, B_j) = \frac{\text{Area}(\text{buffer}(A_i \cap B_j))}{\text{Area}(\text{buffer}(A_i \cup B_j))} \quad (1)$$

$$E(A_i, B_j) = S(A_i, B_j)\omega_1 + P(A_i, B_j)\omega_2 \quad (2)$$

新旧要素的关联度计算方法如公式 (2) 所示。 $E(A_i, B_j)$ 为新旧要素的关联度, $S(A_i, B_j)$ 为新旧要素的属性信息语义关联度, 由新旧要素的属性字段内容对比得到^[15], $P(A_i, B_j)$ 为新旧要素空间关联度, ω_1, ω_2 分别为空间关联度及语义关联度的权重, 且 $\omega_1 + \omega_2 = 1$, 实验发现当取 $\omega_1 = \omega_2 = 0.5$ 时, 检测的效果比较理想。如 $E(A_i, B_j)$ 高于给定阈值, 说明新旧要素之间存在关联关系, 新要素为旧要素变化而来; 否则, 说明新旧要素间不存在关联, 新要素是在旧要素消失后产生。

通过变化检测得到地理要素的变化类型有新增 (新旧要素 0:1)、删除 (新旧要素 1:0)、合并 (新旧要素 $n:1$)、分裂 (新旧要素 1: n)、聚合 (新旧要素 $m:n$)、几何形状与位置变化 (新旧要素 1:1)、属性变化 (新旧要素 1:1) 几种。

1.3 更新方法

基于要素的矢量数据更新是在数据库中将新要素替换旧要素, 更新信息的表达式为:

$$\text{Updateinfo} = \{ \text{OID}, \text{Spatialinfo}, \text{Attributeinfo}, \text{MeasuredTime}, \text{Updatetime} \}$$

其中 OID 为要素的唯一标识, OID 可以是按相关标准编码, 也可以是系统自动生成的 Objectid 或 GUID。Spatialinfo 是空间信息, Attributeinfo 为属性信息, MeasuredTime 是测量时间, Updatetime 是更新时间。由于测量时间与更新时间存在时间差, 测量时间作为更接近于地物的实际变化时间。

更新操作对于新增和删除要素，直接在数据库中插入和删除信息，对于空间信息与属性信息发生改变的要素，先将旧要素删除，再插入新要素。更新过程中需要将变化信息存入变化过程库，同时将删除的要素存入历史库。

1.4 变化信息存储

数据更新引起矢量数据的变化信息需要存储到变化过程库，变化信息的表达式为：

$$\text{ChangeInfo} = \{ \text{ChangeID}, \text{Old} (O_1, O_2 \dots O_m), \text{ChangeType}, \text{New} (O_1, O_2 \dots O_n), \text{ChangeTime} \}$$

ChangeID 为变化信息编号，Old ($O_1, O_2 \dots O_m$) 为变化前旧要素的集合，New ($O_1, O_2 \dots O_n$) 为变化后新要素的集合。ChangeType 为变化类型，变化过程库中记录的要素变化类型如表 1 所示，ChangeTime 为变化时间。

表 1 要素变化类型
Table 1 Changes of type

要素类型	变化类型
点	新增、删除、位置变化、属性变化
线	新增、删除、合并、分裂、聚合、几何形变（延长、缩短）、位置变化、属性变化
面	新增、删除、合并、分裂、聚合、几何形变（扩大、缩小）、位置变化、属性变化

为方便数据表存储，变化过程数据表设计为：

ChangeID	OldOID	ChangeType	NewOID	ChangeTime
----------	--------	------------	--------	------------

OldOID 为旧要素的标识（与更新时存入历史库的要素标识一致），NewOID 为新要素的标识。由于数据更新会引起空间信息和属性信息的变化，因此，一组新旧要素可以同时产生多条变化信息记录。

因要素发生合并、分裂、聚合等变化时，新要素 ID 会发生变化，变化信息表可以起到现状数据与历史数据、不同阶段历史数据之间的要素关联作用。

为了解要素的变化过程，不仅需要了解要素发生了哪些变化，还需要了解要素变化了多少，如地物长度变化的大小，面积变化的大小，哪些属性发生了变化等。地物长度及面积的变化可以通过与历史库中的要素长度及面积字段的对比方便的获取，但对于属性信息的变化，由于各图层属性字段个数不一致，通过与历史库要素信息进行比较需要遍历

所有字段，开销较大，而且属性信息变化表达不够直观。本文设计了属性信息变化存储方式：

$$\text{Attributeinfo} = \{ \text{ChangeID}, \text{Changefields} (\text{field}_1, \text{field}_2, \dots, \text{field}_n), \text{Ovalues} (\text{value}_1, \text{value}_2, \dots, \text{value}_n), \text{Nvalues} (\text{value}_1, \text{value}_2, \dots, \text{value}_n) \}$$

其中，Changefields ($\text{field}_1, \text{field}_2, \dots, \text{field}_n$) 为变化属性字段的集合，Ovalues ($\text{value}_1, \text{value}_2, \dots, \text{value}_n$) 为要素变化前属性字段值的集合，Nvalues ($\text{value}_1, \text{value}_2, \dots, \text{value}_n$) 为变化后要素属性字段值的集合。为方便数据表存储，属性信息变化数据表设计为：

ID	ChangeID	Changefield	Oldvalue	Newvalue
----	----------	-------------	----------	----------

其中 ChangeID 与变化过程数据表中的 ChangeID 关联，Changefield 为发生变化的属性字段名，Oldvalue 为变化前属性字段的值，Newvalue 为变化后属性字段的值。

1.5 历史数据的存储

更新过程中被删除和修改的数据将存入历史库，为方便历史数据的检索，历史数据库的构建参考了时空数据模型，历史信息表达式为：

$$\text{HistoryInfo} = \{ \text{OID}, \text{Spatialinfo}, \text{Attributeinfo}, \text{StartTime}, \text{EndTime}, \text{UpdateTime} \}$$

其中 Spatialinfo 为空间信息，Attributeinfo 为属性信息，StartTime, EndTime 代表该历史数据存在的开始时间和结束时间，UpdateTime 为被更新时间。

变化过程库、现势库、历史库的关系如图 2 所示。

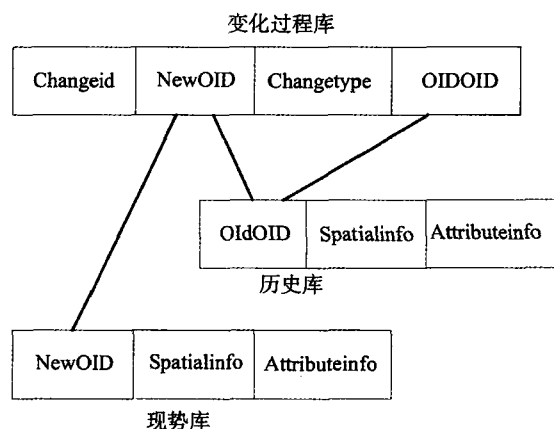


图 2 变化过程库、现势库、历史库之间的关系
Fig. 2 The relationship between change process database, current database and history database

2 要素变化过程信息的应用

2.1 地物生命周期查询

通过变化过程库的查询, 可以了解地理要素在一段时期内的变化情况, 包括地理要素发生了哪些变化, 变化前后地理要素的对应关系。如要了解现状库中某地理要素的变化情况, 可对通过要素编号 OID 对变化过程表进行查询, 能够查询到要素的变化类型, 由哪一个要素变化而来, 通过递归查询, 将上一个状态的要素编号 (OLDROID) 作为新要素变化 (NewOID) 进行查询, 可以得到地理要素的全部变化信息。查询算法伪代码如下:

```
SelectFeatureInfo ( string FeatureID ) // FeatureID 为地理要素的 ID 号
{
    Featureclass tb = Searchchangeinfo ( FeatureID ) //查询变化信息库中 NewOID = FeatureID 的要素信息
    IF ( tb ! = null ) //判断查询结果是否为空
    {
        for ( i = 0 ; i < tb. count ; i + + )
        {
            Saveinfo ( tb. row ) ; //将变化信息存入查询结果队列
            SelectFeatureInfo ( tb [ OldID ] ) ; //递归调用, 将下一状态的 OldID 作为上一状态的 NewID 传入函数
        }
    }
    Else
    Outputinfo ( ) ; //完成查询后输出结果。
}
```

下面以一面要素 A 的变化过程为例说明。在 2008 年 10 月至 2010 年 9 月两年时间内, 面要素 A 发生了面积扩大、名称改名、与其它要素合并 3 次变化。

序号	变化类型	变化时间	变化语义描述
1	面积扩大	20081009	A 面积扩大
2	属性变化	20090105	A 名称改变
3	合并	20100908	A 与 B 合并形成素 C

2.2 线要素长度变化信息

弧段长度是线要素比较关心的信息, 如道路、水系在一段时期内的长度变化。由于变化过程库记

录了变化前后的要素对应关系, 因此很容易对单个要素或多个要素的长度变化进行计算。

单个线要素长度变化为新要素与旧要素的长度差。多个线要素的长度变化计算公式 (3) 所示, m 为要素变化前的数目, n 为要素变化后的数目。

$$\Delta L = \sum_{i=1}^n (\text{len}(\text{NewFeature}_i)) - \sum_{j=1}^m (\text{len}(\text{OldFeature}_j)) \quad (3)$$

L 能够反映线状地物在一段时间内的总体变化趋势, 当 $L > 0$, 说明总体变化趋势为延长, $L < 0$, 说明线总体变化趋势为缩短。

2.3 面要素面积变化信息

面积变化是面要素比较更新的变化因素, 如居民地、植被覆盖面在一段时期内的面积变化大小。通过变化过程库中的新旧地理要素对应关系, 可以方便的对单个或多个面要素的面积变化进行计算。

单个面要素面积变化为新要素与旧要素的面积差。多个面要素的面积变化计算公式 (4) 所示, m 为变化前要素的数目, n 为变化后要素的数目。

$$\Delta S = \sum_{i=1}^n (\text{Area}(\text{NewFeature}_i)) - \sum_{j=1}^m (\text{Area}(\text{OldFeature}_j)) \quad (4)$$

如在 T_1 到 T_2 时间内 ($T_1 < T_2$) 面积发生了多次变化, T_1 到 T_2 时间内面积总变化无须对每次变化进行计算, 可利用历史库中要素的生命周期时间 (StartTime, EndTime), 取时间段内要素的最接近且小于 T_2 时间状态的面积与最接近且大于 T_1 时间状态的面积之差。

当 $\Delta S > 0$, 说明面要素的总体变化趋势为扩张, $\Delta S < 0$, 说明面要素的总体发展变化为缩小。

3 试验案例

本文采用 Visual Studio 2010 结合 ArcGis Engine 10 开发了一套矢量数据增量更新系统, 以某市地形图更新为例来说明系统的应用效果。当采用 1:2 000 地形图数据进行了应用试验时, 在居民地的更新过程中, 系统对新图层中 495 个要素与旧图层的进行了正反向变化信息检测, 发现新增要素 6 个, 消失要素 4 个, 分裂要素 2 个, 合并要素 1 个, 扩大 3 个, 缩小 2 个, 属性变化 16 个。要素的变化信息及新旧要素的对应关系都保存到了变化过程库里的变化过程表, 如表 2 所示。

表 2 变化过程记录表
Table 2 Change process records

变化 ID	旧要素 ID	变化类型	新要素 ID	变化时间
12052	4557	分裂	4622	20120706
12053	4557	分裂	4623	20120706
12054		新增	4625	20120706
12055	4507	删除		20121213
12056	4470	扩大	4624	20121213
12057	4380	合并	4627	20130105
12058	4381	合并	4627	20130105
...

如图 3 所示，在地形图增量更新试验中，以新要素替换了发生变化的旧要素，并对地物的变化信息进行了存储。如图 4 所示，通过对变化信息库进行递归查询，直观准确的了解到编号为 5230 的要素在 2011 年 5 到 2013 年 3 月这段时期内经历了新增、缩小、分裂 3 次变化。

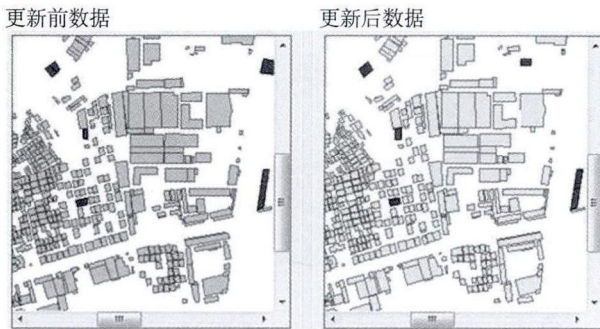


图 3 数据更新实验

Fig. 3 An experiment Vector data updating

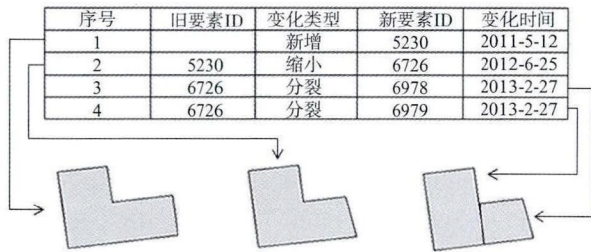


图 4 要素生命周期记录

Fig. 4 Elements of life cycle

试验案例显示，对比以往的矢量数据更新方法^[2,7]，本文提出的方法对更新要素的变化信息进行了识别与存储，有利于历史数据的回溯及地物生命周期的跟踪，还能为地物的变化趋势预测提供参考。

4 结 论

本文采用基于要素的矢量数据更新方式，通过对要素的变化信息进行检测，只更新变化了的要素，实现了真正意义上的增量更新；为顾及地理要素的变化过程，在增量更新的设计中加入了变化过程库，通过变化过程库对更新前后地理要素的变化信息进行了记录，能够建立新旧要素及历史要素之间的对应关系，减轻了历史数据回溯计算量，方便空间数据的管理与维护。

本文设计的更新方法目前只能用于同级比例尺的矢量数据更新，在接下来的工作中将研究多尺度的顾及要素变化过程的增量更新方法。进一步研究包括：多尺度矢量数据的变化信息匹配；多尺度要素变化信息的数据存储结构。

参考文献：

- [1] 陈军,周晓光. 基于拓扑联动的增量更新方法研究[J]. 测绘学报,2008,37(3):322-329.
- [2] 张新长,郭泰圣,唐铁. 一种自适应的矢量数据增量更新方法研究[J]. 测绘学报,2012,41(4):613-619.
- [3] 林艳,刘万增,韩刚. 基态修正的 GIS 数据库增量更新建模[J]. 测绘科学,2012,37(4):199-201.
- [4] 张丰,刘南,刘仁义,等. 面向对象的地籍时空过程表达与数据更新模型研究[J]. 测绘学报,2010,39(3):303-309.
- [5] 薛存金,周成虎,苏奋振. 等. 面向过程的时空数据模型研究[J]. 测绘学报,2010,39(1):95-101.
- [6] 周晓光,陈军,朱建军,等. 基于事件的时空数据库增量更新[J]. 中国图象图形学报,2006,11(10):1431-1438.
- [7] 应申,李霖,刘万增,等. 版本数据库中基于目标匹配的变化信息提取与数据更新[J]. 武汉大学学报:信息科学版,2009,34(6):752-755.

(下转第 141 页)

地结合,如在用水合作博弈中,如何更合理地确定用水联盟的机会收益和核心调节值 ε 等,尚需在今后做进一步研究。

参考文献:

- [1] 郭磊,黄本胜,邱静,等. 核电站淡水用水特征综合分析研究[J]. 水利学报, 2013, 44(5): 615-621.
- [2] 程孟孟,陈进. 南北方典型省份产水和用水特点比较[J]. 长江科学院院报, 2012, 29(11): 1-4, 15.
- [3] 杨有德,张有芷. 长江流域用水状况分析[J]. 人民长江, 2006, 37(8): 10-11, 17.
- [4] 张士锋,贾绍凤. 黄河流域近期用水特点与趋势分析[J]. 资源科学, 2002, 24(2): 1-5.
- [5] 汪党献,王浩,倪红珍,等. 国民经济行业用水特性分析与评价[J]. 水利学报, 2005, 36(2): 167-173.
- [6] LUIS SP, IAN C, IACOVOS I. Improved indicators of water use performance and productivity for sustainable water conservation and saving. *Agricultural Water Management*, 2012, 108: 39-51.
- [7] NEAL J S, FULKERSON W J, HACKER R B. Differences in water use efficiency among annual forages used by the dairy industry under optimum and deficit irrigation[J]. *Agricultural Water Management*, 2011, 98: 759-774.
- [8] 许士国,吕素冰,刘建卫,等. 白城地区用水结构演变与用水效益分析[J]. 水电能源科学, 2010, 30(4): 106-108, 214.
- [9] 张薇,邹志红,王惠文. 城市日用水量预测模型及其应用[J]. 系统工程, 2010, 28(3): 93-97.
- [10] AIREZA G, SAEID E, ALI M, et al. Water transfer as a solution to water shortage: A fix that can Backfire[J]. *Journal of Hydrology*, 2013, 49: 23-39.
- [11] 袁宝招,陆桂华,李原园,等. 水资源需求驱动因素分析[J]. 水科学进展, 2007, 18(3): 405-409.
- [12] 汪贤裕,肖玉明. 博弈论及其应用[M]. 北京: 科学出版社, 2008.
- [13] SALAZAR R, SZIDORSZKY F, COPPLOA E, et al. Application of game theory for a groundwater conflict in Mexico[J]. *Journal of Environmental Management*, 2007, 84: 560-571.
- [14] KAVEH M. Game theory and water resources[J]. *Journal of Hydrology*, 2010, 381: 225-238.
- [15] 沈坤荣,付文林. 税收竞争、地区博弈及其增长绩效[J]. 经济研究, 2006(6): 16-26.
- [16] 陆杉. 农产品供应链成员信任机制的建立与完善——基于博弈理论的分析[J]. 管理世界, 2012(7): 172-173.
- [17] 卢周来. 合作博弈框架下企业内部权力的分配[J]. 经济研究, 2009(12): 106-118.
- [18] 陈晓宏,刘德地,刘丙军,等. 湿润区变化环境下的水资源优化配置——理论方法与东江流域应用实践[M]. 北京: 中国水利水电出版社, 2011.
- [19] 蔡国英,徐中民. 黑河流域中游地区国民经济用水投入产出分析——以张掖市为例[J]. 冰川冻土, 2013, 35(3): 770-775.
- [20] 秦长海,赵勇. 不同节水措施下宁夏水资源利用的经济效益分析[J]. 资源科学, 2008, 30(8): 1135-1140.
- [21] 严婷婷,贾绍凤. 关于国民经济用水投入产出分析指标的讨论[J]. 水利经济, 2009, 27(6): 24-28.

(上接第135页)

- [8] 朱华吉. 基于多级模式的地形数据库变化信息发布方法[J]. 辽宁工程技术大学学报, 2006, 25(1): 32-35.
- [9] 傅仲良,吴建华. 多比例尺空间数据库更新技术研究[J]. 武汉大学学报: 信息科学版, 2007, 32(12): 1115-1118.
- [10] COBB M A, CHUNG M J, FOLEY III H, et al. A rule-based approach for the conflation of attributed vector data[J]. *GeoInformatica*, 1998, 2(1): 7-35.
- [11] KIELER B, HUANG W, HAUNERT J H, et al. Matching river datasets of different scales[M]//Advances in GIScience. Springer Berlin Heidelberg, 2009: 135-154.
- [12] DEN M, LI Z L, CHEN X Y. Extended Hausdorff distance for spatial objects in GIS[J]. *International Journal of Geographical Information Science*, 2007, 21(4): 459-475.
- [13] WALTER V, FRITSCH D. Matching spatial data sets: a statistical approach[J]. *International Journal of Geographical Information Science*, 1999, 13(5): 445-473.
- [14] 童小华,邓慷慷,史文中. 基于概率的地图实体匹配方法[J]. 测绘学报, 2007, 36(2): 210-217.
- [15] 王育红,牛亚辉,林艳. 顾及语义差异的基础地理信息客户数据库更新实施模型[J]. 地理与地理信息科学, 2011, 27(1): 1-6.