

土地利用数据多尺度表达规则提取与应用

高文秀 侯建光 朱俊杰

(武汉大学测绘遥感信息工程国家重点实验室, 武汉 430079)

摘要 土地利用数据综合是从大型土地利用数据库中派生多尺度数据的基本操作,提高土地利用数据综合的客观性和可靠性的有效途径是解析数据自身特征、从数据特征导出影响综合操作的规则、遵循规则实施土地利用数据综合操作。本文主要探讨随地理区域不同而变化的区域依赖性数据综合规则的提取和应用。首先阐明了数据综合规则的实现目标和规则层次性;然后详细描述了利用空间关联关系分析和景观指数分析来确定用地类型的重要性级别、各用地类型地块的最小面积域值和空间分布状态等规则的提取方法和应用,并以湖北省九峰县1:10 000土地利用数据综合生成1:50 000数据为例检验上述规则的有效性。

关键词 多尺度土地利用数据 数据综合 数据综合规则 空间关联分析 景观指数

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2009)06-1024-06

Extraction and Application of Multiple Representation Rules for Land-use Data

GAO Wen-xiu, HOU Jian-guang, ZHU Jun-jie

(State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079)

Abstract Land-use data generalization is one of the basic operations for deriving multi-scale data from a comprehensive land-use database. Acquiring data characteristics and converting them into generalization rules used to implement data generalization is an effective way to improve the objectivity and validity for land-use data generalization. This paper focuses on the extraction and application of region-dependent rules for land-use data generalization. First, the objectives and hierarchy of the rules are clarified. Then, spatial association analysis and landscape indices are employed to set the importance ranks, minimum area thresholds and spatial distribution pattern of each land-use class. Finally, a case study was undertaken to show the application of these rules using 1:10 000 land-use data around Jiufeng county of Hubei, China.

Keywords multi-scale land-use data, data generalization, generalization rules, spatial association analysis, landscape index

1 引言

随着数据库技术和地理信息系统技术的发展,土地利用数据的存储和管理大都转向了数据文件或大型数据库方式。在实际应用中,常常需要依据不同应用需求和地理区域范围派生不同尺度的土地利

用数据。空间数据综合是实现多尺度地理空间数据表达的重要途径^[1],同样也是从数据库中派生多尺度土地利用数据的有效手段。目前多采用以人机交互为主的数据综合方式^[2],很大程度上是对手工数据综合过程的模拟,对人脑的主观判断具有高度依赖性^[3],其客观性、可靠性和有效性有待于大力提高。改善这个过程的有效途径是从数据自身特征导

基金项目:国家重点基础研究发展计划(973)项目(2006CB701304);国家自然科学基金项目(60601082)

收稿日期:2008-12-03;改回日期:2009-03-15

第一作者简介:高文秀(1971~),女,副研究员。2002年于武汉大学获摄影测量与遥感专业博士学位。主要研究方向为地图数据综合、地理信息标准化等。E-mail:wxgao@lmars.whu.edu.cn

出影响综合操作的规则,基于规则决策数据综合操作的顺序和综合算子的选择^[4]。

有些影响综合操作的规则是通用的,普遍适用于不同区域、不同尺度的土地利用数据的综合和表达,例如保持土地利用类型表达的逻辑一致性,同类用地的父类和子类不能同时出现在同一尺度的数据中^[5]。而有些规则则依据地理区域的不同而有所变化,例如不同地理区域各种用地类型的重要性不同,城市区域以表达城市居民地、商业用地、交通用地等为主,农业区域以表达耕地、水域、经济林地为主;此外,随着地理区域的不同,不同用地类型之间的地理分布依赖关系也不同,各种尺度的土地利用数据都要能够正确反映土地利用的地域差异。本文将这类规则称为区域依赖性规则。

前面所说的通用规则属于规范性规则,不会随着地理区域的变化而变化,在多尺度数据提取过程中可以直接运用这类规则;而区域依赖性规则则要从实际的土地利用数据集中通过认识和解析数据特征适时确定。本文首先分析了区域依赖性规则的实现目标和层次特征,然后借助于空间关联关系分析方法和景观指数分析方法提取相关规则,并阐述这些规则在数据综合过程中的应用。

2 区域依赖性规则

2.1 规则的层次特征

认识数据首先要分析数据所表达的地理现象的分布特征。整体上讲,土地利用属于连续分布的地理现象,在地图上土地利用地块(表现为线状或面状图斑)布满了整个制图区域;但是从单个用地类型的角度讲,同一地类的地块是离散分布的,某些地类地块均匀分布而某些地类地块聚集分布。因此,土地利用数据的表达要同时顾及土地利用整体连续分布和单个地类地块分散分布的特征,这也是区域依赖性规则所要实现的目标之一。

根据规则约束的对象层次,规则又可细分为地理区域层面的规则、用地类型层面的规则和用地地块层面的规则。地理区域层面的规则以整个地理区域为约束对象,如地理区域范围内各种用地的重要性级别、不同用地类型地块的地理分布关系和各类用地的面积比率等;用地类型层面的规则是针对某个特定用地类型的规则,如某用地类型地块的最小面积阈值,用地类型不同其最小面积阈值不同;用地

地块层面的规则用于约束某个或某几个具体用地地块的表达,例如一组邻近但不邻接且形状相似的鱼塘或池塘,在不能清晰地描述所有对象图斑时,需要通过聚合、移位、放大等操作将其组合成数量减少了的图斑,既要反映成组分布的模式,又要考虑可读性。

2.2 区域性规则的提取方法

本文利用空间关联关系分析方法和景观指数分析方法分析设定用地类型的重要性级别、各用地类型地块的最小面积阈值和空间分布状态三个方面的规则。

(1) 用地类型的重要性级别

设定用地类型重要性级别的目的在于区分指定地理区域的主导用地类型和次要用地类型,以便于在数据综合过程中,突出反映主导用地类型的地块分布特征。

判定用地类型重要性级别的最基本条件是所有用地类型的地块面积百分比,通常占地面积越大,用地类型重要性级别越高(不包括未利用地)。在此基础上,依据其他用地类型地块与该用地类型地块在空间分布上的关联关系,调整各用地类型的重要性级别。例如,如果某类面积不占优,但其地块与重要性级别最高的用地地块在空间上存在很强的相邻分布关联关系,则一定程度上说明两种用地类型之间存在用途上的依赖关系,那么该面积不占优的用地类型的重要性级别应该提高。为此,利用空间关联关系分析方法提取不同用地类型地块在空间分布上的相邻分布关联关系,然后基于式(1)算出各类用地重要性指数。

$$limpi = A_i + \sum_{j=1}^n (A_j \omega_j) \quad (1)$$

式中, $limpi$ 为第*i*类用地类型的重要性指数, A_i 为第*i*类用地类型占地面积百分比, n 为用地类型的个数, ω_j 为第*i*类用地类型地块与第*j*类用地类型地块之间在空间上相邻分布的关联关系置信度,本文称其为关联强度,按照式(2)计算。

$$\omega_j = count(T_i - > T_j) / count(T_j) \quad (2)$$

式中, $count(T_j)$ 表示属于第*j*类用地类型的地块数目, $count(T_i - > T_j)$ 表示第*j*类用地类型地块中与第*i*类用地类型地块在空间相邻分布的地块数目。这其中需要排除地块数目较少的用地类型,因为其支持度较低(即该用地类型地块数目占整个地理区域地块总数的百分比很小)。这类用地将被直接归

属到次要用地级别。

(2) 各用地类型地块的最小面积阈值

设定最小面积阈值的目的在于消除那些无法在特定尺度上清晰表达或者不需要详细表达的次要地块,一方面可以增强数据表达的可读性,另一方面概化非重要用地类型的地块分布特征。因此,通常重要用地类型的地块最小面积阈值略小,而非重要用地类型的地块最小面积阈值则略大。

本文基于各类用地的重要性指数来确定各类用地的最小面积阈值。首先将重要性指数根据其自然分布状况进行分级,从最重要级别到最次要级别依次用数字1,2,⋯,R表示。然后采用经验式(3)为各个级别的用地类型设置最小面积域值。

$$A_{\min_i} = 1 + MI_i \quad (3)$$

式中,

$$I_i = \frac{r_i - 1}{R - 1} \quad (4)$$

式中, A_{\min_i} 为第*i*类用地类型地块的最小面积域值, r_i 为第*i*类用地类型所属级别对应的数值,为了消除级别内的差异,利用式(4)将 r_i 标准化为 I_i , M 为常数,是根据具体应用需求设定的经验值。

(3) 地块的空间分布状态

认识数据描述的土地利用空间分布状态是在某一尺度上正确表达土地利用分布特征的基础。根据前述规则的层次性,需要了解用地类型层面和用地地块层面的空间分布状态。前者用于在综合过程中保持各类用地地块的空间分布差异,后者用于保持同类用地在局部区域的特殊分布特征。本文主要采用两个景观指数分别从用地类型层面和地块层面上分析各类图斑的空间分布特征:斑块密度(patch density)和香农均匀度指数(Shannon's evenness index)。斑块密度是对各类地块分布特征的最直观表示,其值为研究区域内每类用地的地块个数与区域总面积之比,香农均匀度在本文中用于反映一种地类的土地面积在其所有地块上分布的均匀程度,其值越小说明其分布越不均匀^[6]。

3 区域依赖性规则的应用

土地利用数据综合的目的简单来讲就是在特定尺度上尽可能地保留和反映重要用地类型地块的空间分布特征。从另外一个角度讲,就是尽可能概化非重要用地类型地块特征以突出重要用地类型特

征。因此,通常土地利用数据综合都是从存在空间冲突的、非重要的问题图斑开始,逐个逐步地消除问题图斑对清晰地表达整个土地利用特征的干扰,同时概化次要用地类型特征,保留并突出重要用地类型的特征。基本步骤如下:

(1) 根据各类用地的最小面积阈值搜索所有小于阈值的地块图斑,在本文的后续部分将这类图斑统称为小图斑。

(2) 依次处理次要用地类型的小图斑,遵循的基本原则是首先考虑与其相邻的同父类图斑进行合并,以保持其大类(即父类)层面上用地的面积比率;其次对于岛屿类图斑考虑与其背景图斑融合,忽略次要的小图斑;第三考虑将小图斑进行压缩并将面积分摊到它的邻接图斑中。这一步骤一方面消除了次要地类的小图斑,另一方面与它们邻接的重要地类的小图斑可能因获取了该次要小图斑的面积而不再是小图斑,或者改善邻接图斑局部范围的几何形状,这样即保留了重要用地地块,同时也节省了后续的数据处理工作。

(3) 对于重要用地类型的小图斑,首先考虑与其邻接的属于相同父类的图斑进行合并;对于岛屿类图斑则要分析搜索其周围一定范围内是否有同类图斑存在,如果有则采用移位和聚合相结合的操作或者放大操作,即消除了小图斑问题,又保持该类地块在这个局部空间上的分布特征;对于一定范围内没有同类型图斑的孤立小图斑,依据其邻接图斑用地类型和该小图斑用地类型之间的关联强度做出相应的处理:如果存在强关联的地块,则要扩大该孤立图斑,尽量保持两类用地之间的关联关系;如果不存在则可以考虑与其背景图斑融合。

4 应用示例

实验数据是湖北省九峰县2002年的1:10 000土地利用数据,该数据按照中国土地利用三级分类体系的第三级类型详细记录了各地块的土地类型和相关属性信息,如权属性质、权属单位、坡度等。试验区域属于农业用地区域,其覆盖总面积为2 555 m²。

4.1 区域依赖性规则的提取与分析

利用前面描述的区域依赖性规则提取方法,分析设定该实验区域用地类型的重要性级别、最小面积域值和空间分布状态,并用于指导生成1:50 000土地利用图的数据综合过程。

(1) 用地类型重要性级别的设定

图 1 显示了各类用地的面积比率,其中用地类型为灌溉水田(编码为 111)的面积比率最大,在忽略其他影响因素(如各类用的经济收入等)的前提下,假定 111 是该区域的主要用地类型,重要性级别最高。图 2 显示了其他各类用地地块与 111 用地类型地块的空间相邻分布的关联关系强度,其中排除了支持度小于 0.5% 的用地类型。图 3 是依据式(1)计算得到的各类用地的重要性指数。

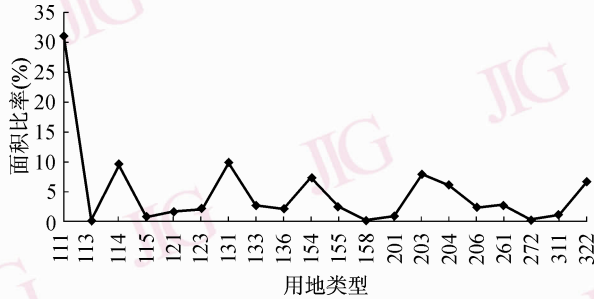


图 1 各类用地的面积比率
Fig. 1 Area ratio of all landuse classes

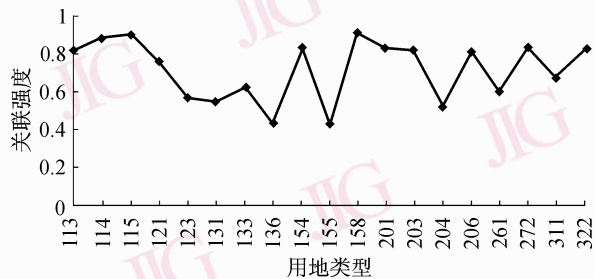


图 2 各类用地地块与 111 用地类型地块的空间相邻分布的关联关系强度
Fig. 2 Strength of association relation on spatial adjacency between landuse parcels of Class 111 and other classes

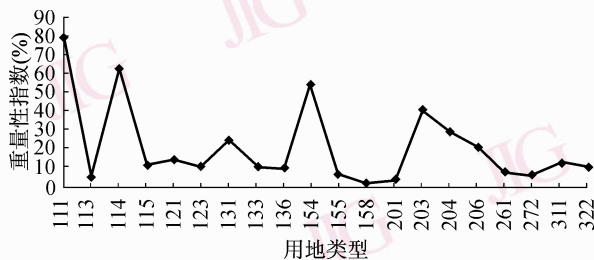


图 3 各类用地的重要性指数
Fig. 3 Importance index of all landuse classes

可以看出某些面积并不占优的用地类型如坑塘水面 154(小于 10%),由于其地块分布与用地类型

111 地块有较强的空间相邻分布关联关系(大于 0.8),经过式(1)的调节之后,其重要性提高,这也符合实际的土地利用分布基本常识,即灌溉水田对水源的依赖性。类似的还有旱地 114、农村居民地 203 等用地类型。而水浇地 113 尽管与 111 地块有较强的关联关系,但是由于占地面积过小,因而最终重要性级别较低。

为了简化处理过程,对重要性指数进行分级。根据其自然分布状况,将重要性指数大于 50% 的用地类型设置为强重要性级别,赋值为 1;20% ~ 50% 之间设置为中等重要性,赋值为 2;小于 20% 设置为非重要性,赋值为 3。根据实际应用需求和数据,重要性级别数可以有所不同。本文实验数据中各类用地重要性级别如表 1 所示。

表 1 各类用地重要性级别及最小面积阈值
Tab. 1 Importance indices and minimum area thresholds of all landuse classes

用地类型	重要性级别	最小面积阈值
111, 114, 154	1	1
203, 204, 131, 206	2	6
121, 311, 115, 322, 123, 133, 136, 261, 155, 272, 113, 201, 158	3	11

(2) 设置各用地类型地块的最小面积阈值

传统的手工综合往往采用固定的最小面积阈值来限定各类用地的上图尺寸,具有一定的普遍性,但是不同区域采用相同阈值可能无法真实反映区域差异。为克服上述不足,基于式(3)计算了各级用地地块最小面积阈值(如表 1 所示),其中式(3)中的常数 M 为 10。

(3) 空间分布状态分析

图 4 和图 5 分别描述了研究区域内各用地类型的地块图斑密度和香农均匀度指数。大部分图斑密度较大的用地类型其图斑面积分布也较为均匀(香农指数较大),如 154, 114 和 203 类型等。154 类型的图斑个数多,香农指数大,但图 1 的面积比率显示 154 的地块面积并不占优,远小于 111 类型,这说明属于 154 类型的图斑尺寸不大,面积分布较为均匀;而 111 类型的占地面积最大,但是图斑密度和香农指数相对较小,说明属于 111 类型的图斑面积分布不均匀,有的图斑面积较大,有的面积较小。这一点与目视判断整个图面的结果非常相符。对于面积均匀分布的用地类型,在图斑综合处理过程中,要顾及

其周围一定范围内是否存在同类图斑,以保持图斑均匀分布的特征。此外,这些指数还可以作为综合结果是否保持了各用地类型,尤其是重要用地类型的整体分布特征判断依据之一。

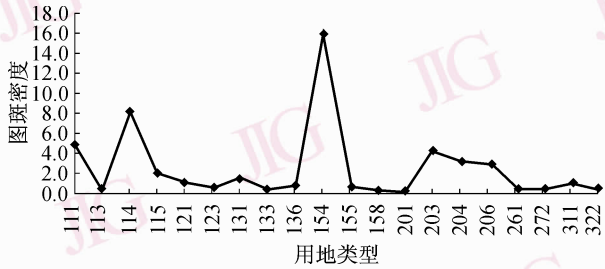


图 4 图斑密度

Fig. 4 Density of landuse patch

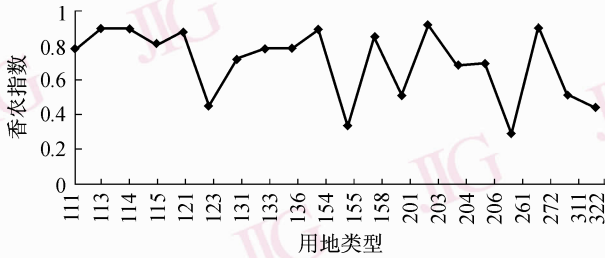


图 5 香农均匀度指数

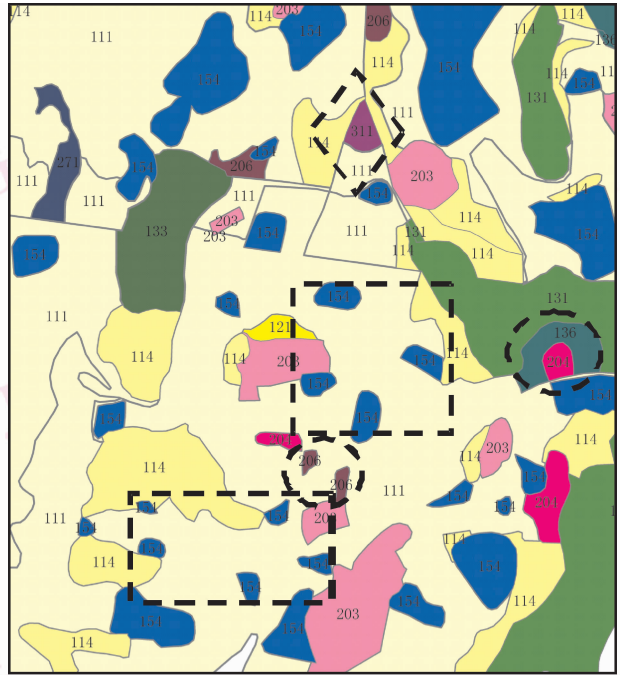
Fig. 5 Shannon's evenness index

4.2 基于区域依赖性规则的土地利用数据综合实验

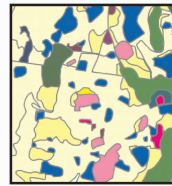
图 6 给出了基于区域性规则的实验数据综合结果。图 6(a) 按比例显示了部分 1 : 10 000 实验数据,图 6(b) 是按比例显示的 1 : 50 000 未经过数据综合的土地利用数据,图 6(c) 是经过数据综合后的 1 : 50 000 数据。

图 6(a) 图斑上标注的数字是各用地类型的编码,图面上的两个用虚线圆圈标定的部分是对次要用地小图斑的处理,采用前面阐述的与相邻的同父类图斑合并的规则,将 136 类型(苗圃)合并到相邻的 131 类型(有林地),归属同一父类 13 类型(林地);206 类型(特殊用地)合并到相邻的 203 类型(农村居民点),归属同一父类 20 类型(居住用地)。

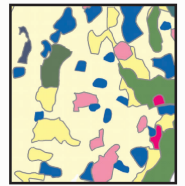
图面上部用虚线菱形框标定部分的小图斑 311 类型(荒草地),由于没有邻接的同父类图斑,因而采用面积分割方法将其面积分配到邻接的 111 和 114 类型图斑中,改善了 114 类型图斑狭长部分的几何图形,节省了后续对该部分的综合操作。对比



(a) 1:10 000原始地图



(b) 1:50 000未经过综合处理的地图



(c) 1:50 000综合处理后的结果

图 6 土地利用图

Fig. 6 Landuse map

图 6(b) 和图 6(c), 该部分的 114 类型图斑,图 6(c) 上 114 类型图斑具有良好的可读性。

图面中部和左下部虚线方框标定的部分 154 类型的小图斑属于重要性级别较高的图斑,在中部方框的 154 类型图斑面积比较接近最小面积图斑,且相互距离比较大,所以采用放大重要级别较高的小图斑规则,在图 6(c) 上即反映了此处 154 类型图斑的分布状态,同时具有良好的可读性。但是在左下部方框内的 154 类型图斑,由于面积远小于最小面积阈值,而且在其周围有较大面积的同类图斑存在,则采用了与背景图斑融合的规则。

除了对小图斑进行处理之外,还对 1 : 10 000 地图上由于坡度不同而区分为不同图斑的相邻 111 类型图斑进行了合并,在 1 : 50 000 地图上忽略了 111 类型地块在坡度上的差异。例如 1 : 10 000 图面上左侧两个上下相邻的 111 类型图斑在 1 : 50 000 图面上被合并为一个图斑。

5 结 论

为了增强从数据库中派生多尺度土地利用数据操作的客观性和可靠性,本文尝试利用空间关联关系分析和景观指数分析方法从原始土地利用数据中解析数据自身特征,并将其转化为数据综合规则,用于指导实现多尺度数据提取的数据综合操作。

提取多尺度土地利用数据的宗旨是在不同尺度层面上反映地理区域的重要用地特征,因此确定用地类型的重要性是关键问题之一。本文联合各类用地的占地面积比率、不同用地类型地块在空间分布上相邻关系、地块数目等多种因素来确定各类用地的重要性级别,不仅顾及了父类用地类型的整体特征,还考虑了子类之间的差异。小图斑的确定是整个数据综合操作的关键步骤之一,与各类用地的重要性级别相结合来设定各类用地的最小面积阈值,相对于手工综合中为不同区域数据设定固定阈值,增强了数据综合操作的适应性和可靠性。空间分布状态的确定为数据综合操作提供了有利依据,一方面可以指导数据综合操作过程的决策,另一方面可以作为评价数据综合结果的依据。从不同层面上分析用地地块的空间分布状态,既能维护地块在地理区域上的整体分布状况,又能适当地保持地块的局部分布特征。

虽然上述规则在一定程度上提高和改善了土地利用数据综合的可靠性和客观性,但是跟手工数据综合相比,数据操作人员可能会运用到比上述规则更加丰富的知识和经验;而且通过空间关联关系分

析发现的有些知识和规则并不一定完全适用于综合操作,还需要在一定程度上借助于人工判断。景观指数能够很好地描述各类用地的分布状态,但是对空间分布状态的解读转化为自动指导综合过程运行的规则还需要进一步的研究和实验。

致 谢 武汉大学吴华意教授对本文用地类型重要性指数公式的确立给予重要指点,杨必胜教授对本文的写作提出了大量修改意见,特此致以衷心感谢!

参考文献 (References)

- 1 Wang Jia-yao, Cheng Yi. The multiple character of spatial data and automated generalization [J]. Hydrographic Surveying and Charting, 2004, **24**(4): 1-3. [王家耀,成毅. 空间数据的多尺度特征与自动综合[J]. 海洋测绘, 2004, **24**(4):1-3.]
- 2 Li Zhi-gang, Zhang Xiao-yong, Ai Ting-hua. Generalization research of land use patch[J]. Geospatial Information, 2004, **2**(3): 13-18. [李志刚,张小勇,艾廷华. 土地利用图斑综合研究[J]. 地理空间信息, 2004, **2**(3):13-18.]
- 3 Wu Fang, Wang Jia-yao. Analysis and research of concept frame of map automatic integration [J]. Engineering of Surveying and Mapping, 2002, **11**(2):18-21. [武方,王家耀. 地图自动综合概念框架分析与研究[J]. 测绘工程, 2002, **11**(2):18-21.]
- 4 Weibel R, Dutton G H. Constraints-based automated map generalization [A]. In: Proceedings of the 8th International Symposium on Spatial Data Handling[C], Vancouver, BC, Canada, 1998: 214-224.
- 5 GAO Wen-xiu, Gong Jian-ya, Li Zhi-lin. Thematic knowledge for the generalization of land use data[J]. The Cartographic Journal, 2004, **41**(3):245-252.
- 6 Christopher M, Bishop. Pattern Recognition and Machine Learning [M]. New York: Spring Press, 2006.