

基于语义优先的土地利用图斑综合的研究 ——以大连旅顺口区北海街道为例

杨 俊^{1,2}, 席建超¹, 孔凡强², 葛全胜¹, 李雪铭², 高 航²

(1.中国科学院地理科学与资源研究所, 北京 100101; 2.自然地理与空间信息科学辽宁省重点实验室, 辽宁 大连 116029)

摘要:土地利用图斑综合是实现其多尺度表达的关键技术, 主要包括相邻图斑的融合、相离且语义邻近图斑的聚合、狭长图斑的综合以及要素边线的化简等。通过以下4种方法实现图斑综合: ①以要素间语义相似度为先决条件并顾及空间拓扑关系建立邻近度分析模型, 通过模型检索要素的最邻近图斑实现数据融合; ②以相离要素缓冲区交集集中的节点建立多边形要素填补桥梁区域实现要素聚合; ③以相邻要素的缓冲区将狭长区域剖分, 并融合到周围要素的方法来实现狭长图斑的综合; ④以Douglas-Peucker算法实现要素边线化简。上述方法对普通图斑及特征图斑均提出了各自适应的综合算法。实验结果表明上述模型及算法能最大限度的保证各地类面积平衡、也能有效控制要素变形、合理压缩数据并保证地图简洁、美观, 保证自动化制图综合的简捷高效。

关键词:语义相似; 制图综合; 边线化简; 土地利用图斑

中图分类号: P283

文献标识码: A

文章编号: 1000-0690(2013)08-0949-08

土地利用数据中要素繁多且相互之间存在着复杂的空间和语义关系, 对其进行综合要保持综合前后土地利用信息最大程度的一致性, 满足该条件且保证数据特征的综合是一项十分困难与复杂的工程, 因此制图综合至今仍然是地图学的重点与难点^[1]。近些年, 随着计算机与GIS技术的快速发展, 国内外许多学者对此做了诸多的研究, 涌现出了许多土地利用数据综合的算法与模型: 有针对点、线、面具体化简方法的探究^[2-10], 多是基于Douglas-Peucker算法和Delaunay三角网生成骨架线的方式来实现; 有针对影像地图制图综合的探究^[11], 主要从比例、目的、视觉3个方面阐述; 有基于化简前后面积平衡的算法探究^[12, 13], 通过改变图斑边界线来保证综合前后面积的绝对平衡, 保持数据的面积一致性; 有针对具体综合方法的研究^[14-21], 如分布式环境下综合方式的实现, 综合指标体系的建立等; 以上成果在各自的研究领域均有不错的适应性。

土地利用图斑数据的综合主要是对矢量面状

数据进行综合, 上述学者在处理斑块时基本都引入三角网对其进行剖分处理, 算法复杂, 对于大量的土地利用数据处理十分缓慢; 在建立综合模型时将要素的语义邻近与空间邻近综合考虑, 通过赋予各部分相应的权值求和得出量化结果, 但是土地利用综合前后对各地类总面积要求严苛, 综合后各要素总面积须在一定幅度内波动, 因此语义临近比空间临近优先级高, 在语义邻近度相同的情况下, 将要素空间拓扑关系作为辅助因素决定最终结果, 这样便可最大限度的保证各地类的总面积。因此本文在传统综合方法的基础上, 引入缓冲区实现对要素聚合与狭长地物的综合处理, 并根据土地利用类型的语义相似度和空间拓扑关系提出要素临近度分析模型, 这对于土地利用数据的快速综合与保持数据综合前后的一致性有十分重要的作用, 同时对于体现要素语义对制图综合的影响以及保证地图的美观与可用性方面均有十分重要的意义。

收稿日期: 2012-12-05; **修订日期:** 2013-05-20

基金项目: 国家自然科学基金项目(40971299, 41171137) 和教育部人文社科项目(09YJC790135)资助。

作者简介: 杨 俊(1978-), 男, 湖北孝昌人, 博士, 副教授, 主要从事区域地表过程、城市人居环境与地理信息系统应用。E-mail: lnu_hsrc@126.com

通讯作者: 席建超, 副研究员。E-mail: xijc@igsrr.ac.cn

1 要素邻近度模型及综合算法

1.1 要素邻近度模型

1.1.1 要素语义邻近度

土地利用数据在空间上全覆盖、无缝隙、无重叠,在语义上具有类型的层次化划分^[14-17],其综合是在以上前提下对要素集进行概括。根据中华人民共和国土地管理行业标准 TD/T1028-2010 乡(镇)土地利用总体规划数据库标准附录 B 中土地利用用途分类,数据分 3 个一级类,二、三级地类依次对其进行细分如图 1 所示(由上至下每一层是一级)。语义的多样性使其综合非常困难,因此只有定义出语义之间的关系,确定语义邻近度,才能制定明确的综合规则。

土地利用数据往往关心的是一、二级各地类的总量,三级地类中仅对城乡建设用地的子类有兴趣,因此本文认为同属于一级地类的要素语义才会存在语义临近,否则认为语义无关。对每个二、三级用地类型制定一个同级的语义临近的地类序列,以三级地类旱地为例,首先要考虑与旱地同父类的地类,这里顺序为旱地、水浇地、水田。然后再考虑非同父类的地类与旱地的临近关系,即水田之后为园地、林地、牧草地、田坎、农田水利用地、设施农用地、农村道路。

1.1.2 模型中要素关系定义

这里设土地利用数据为 $LandUseSet = \{F_1, F_2, F_3 \dots F_n\}$, S_{Area} 、 $D_{Feature}$ 分别表示要素的最小上图面积和要素间的最小距离;要素地类名称用 $LandName(F_n)$ 表示,要素地类所属的父地类(如耕地、园地的父地类是农用地)用 $Father[LandName(F_i)]$ 表示,要素的面积用 $Area(F_i)$ 表示, $Dis(F_i, F_j)$ 表示要素 F_i 与 F_j 的最小距离;要素 F_i 与 F_j 的空间拓扑关系用 $TopoRel(F_i, F_j)$ 来表示,语义相似度用 $SemRel(F_i, F_j)$ 表示。 $TopoRel(F_i, F_j)$ 和 $SemRel(F_i, F_j)$ 的取值如下:

1) $TopoRel(F_i, F_j)$ 的取值为 -1、0、1。首先判断

$CommonLine(F_i, F_j)$ (判断两要素是否共线),若共线说明要素 F_i, F_j 空间邻接,此时 $TopoRel(F_i, F_j) = 0$; 否则判断 $Dis(F_i, F_j)$ 与 $D_{Feature}$ 的关系,若 $Dis(F_i, F_j) < D_{Feature}$ 表示要素 F_i, F_j 空间邻近,此时 $TopoRel(F_i, F_j) = 1$; 否则 $TopoRel(F_i, F_j) = -1$, 表示要素 F_i, F_j 空间无关。

2) $SemRel(F_i, F_j)$ 的取值范围由 F_i 邻近地类的个数而定,如图 1,耕地的临近地类有 10 种(包括自身),当 F_i 为耕地时, $SemRel(F_i, F_j)$ 按与耕地的语义邻近度赋值依次取值为 0、1、2、……9; 当要素 F_i, F_j 语义无关时 $SemRel(F_i, F_j) = -1$ 。

1.1.3 模型规则

土地利用综合前后常常要统计各行政区中地类的面积,因此行政区可以说是一个独立的综合单元,以下规则都是在该前提条件下制定的。综合前要先求出次要要素集 $FeaSet\{F_i\}$ ($Area(F_i) < S_{Area}$)。根据 1.1.2, 当 $TopoRel(F_i, F_j) = -1$ 时, F_j 距离 F_i 太远两者不发生关系,而当 $SemRel(F_i, F_j) = -1$ 时,两要素语义无关,不能对其进行聚合处理。由此模型的处理过程如下:步骤一:根据条件 ① $SemRel(F_i, F_j) = 0$ And $TopoRel(F_i, F_j) = 0$ 检索要素集合 $FeaSet\{\}$, 该集合中与要素 F_i 共享边界最长的就是所求要素。如图 2a 中的要素 F4 为次要要素,满足条件 ① 的集合应该是 $FeaSet\{F_6, F_{13}\}$, 此时集合中的元素与 F4 共享边界最长的就应该是所求的最邻近要素,即为 F6。若满足条件 ① 的集合为空则进行步骤二:根据条件条件 ② $SemRel(F_i, F_j) = 0$ And $TopoRel(F_i, F_j) = 1$ 检索要素集合 $FeaSet\{\}$; 该集合中位于要素 F_i 的 $D_{Feature}$ 半径的缓冲区中面积最大的要素就是我们所求的要素。例如图 2b 中的 F7, 满足条件 ① 的集合为空,而满足条件 ② 的集合为 $FeaSet\{F_1, F_{11}\}$, 集合中含有 2 个元素,就要以 $D_{Feature}$ 为半径作 F7 的缓冲区 $Buffer(F_7)$, 分别看要素 F1 与 F11 位于 $Buffer(F_7)$ 中的面积,由于 F1 位于 $Buffer(F_7)$ 中的面积大,所以 F1 就是所求的要素。经过上述两步仍没有检索出最邻近要素时,依次将 $SemRel(F_i, F_j)$ 的值加 1 再循

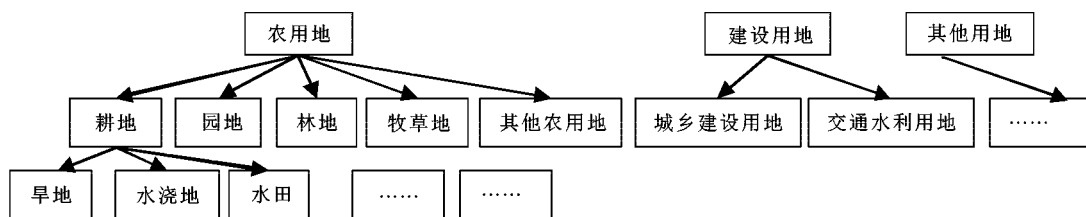


图1 土地类型语义层次树

Fig.1 Hierarchical tree of land use type

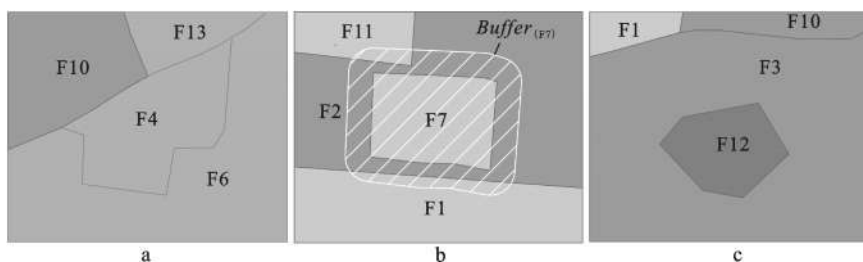


图2 图斑邻近度示例

Fig.2 An example of neighborhood degree

环,直到检索出最邻近要素。若 $SemRel_{(F_i, F_j)}$ 取到最大值仍找不到上述几类元素,就将要素 F_i 融合到邻接最大面积的要素中,如图 2c 中的要素 F12,最终被合并到 F3 中。上述过程中若是最终检索出的要素 F_i 与 F_j 空间邻接,就采用融合方法进行处理,若 F_i 与 F_j 空间邻近,就采用聚合方法进行处理。该模型以要素语义邻近为主,空间拓扑关系为辅来判断次要要素的最邻近要素,模型过程简洁并能够尽量减少综合过程中各个地类的面积变化幅度,满足土地利用综合的要求。本邻近分析模型的流程图如图 3 所示。

1.2 要素处理算法

1.2.1 聚合处理

要素聚合是空间上相离要素的合并,采用要素聚合能防止同类距离较近的元素被删除,防止综合过后该地类的面积总量变化过大^[18]。本文采用缓冲区聚合算法的具体步骤如下:以图 4a 中要素 F1 为例:① 缓冲半径采用 $D_{Feature}$ (要素间最小距离)对 F1 做缓冲区 $Buffer_{(F1)}$;② 寻找与 $Buffer_{(F1)}$ 相交的同类图斑 F2;③ 以 $D_{Feature}$ 为缓冲半径做 F2 的缓冲区 $Buffer_{(F2)}$,如图 4a 所示;④ 求 2 个缓冲区的交集 $Buffer_{(F1)} \cap Buffer_{(F2)}$,图 4a 中两要素的缓冲区交集为中间网格区域部分;⑤ 求出 F1、F2 两要素在缓冲区交集(图 4b 中黑色边界)中的节点集合 $NodeSet \{N1, N2, N3 \cdots Ni\}$ 如图 4b 所示;⑥ 以 $NodeSet$ 集合中的节点建立多边形要素 F_n ,即为图 4b 中中间黑色阴影区域;⑦ 将要素 F1、F2、 F_n 合并即可生成聚合后的要素如图 4c 所示。

由图 4b、c 可以看出,利用缓冲区交集中节点生成的要素正是相离要素的桥梁区域,再将该区域与原要素合并便可有效的消除桥梁,使相离要素合并为一起,且有效的保持了要素的原有形状特征,达到了综合的要求。用该方法进行聚合处

理时要注意要素重叠的处理,对于新加的要素与桥梁区域的重叠直接将桥梁区域切除掉即可。

1.2.2 狭长要素处理

土地利用数据中的狭长要素主要包括铁路、公路、田坎、沟渠等,对其处理时不能简单的与周围要素融合或者聚合,因为该数据长而窄,综合时仅仅将其进行简单合并处理并不能消除该要素对数据整体的影响。本文提出一种基于缓冲区的方法对狭长要素按语义相似度进行分割,并将分割后的要素融合到周围的要素中。该方法算法简单容易实现且效率高。

考虑到狭长要素与周围两侧空间邻接要素的语义相似度,先提取出狭长地物的中心线(如图 5a 中的黑色线 CenterLine),用中心线将狭长要素分为左右 2 部分(图 5a 中的 Road (Left)、Road (Right)),待分割要素 Road (Left) 只用与之空间邻接的要素 F1 来分割,虽然 F2 和 F3 也与道路 Road 空间邻接,但是他们不直接与 Road (Left) 接触,不参与其分割,而右侧待分割要素 Road (Right) 与 F2、F3 空间邻接,因此用要素 F2、F3 对其分割。现在以分割 Road (Right) 为例说明对狭长要素处理步骤:① 建立与 Road (Right) 空间邻接要素 F2、F3 的缓冲区(缓冲距离采用狭长地物最宽长度的二分之一) $Buffer_{(F2)}$ 、 $Buffer_{(F3)}$,缓冲区之间必定存在重叠,如图 5b 所示;② 根据要素 F2、F3 与 Road 的语义相似度 $SemRel_{(F2, Road)}$ 、 $SemRel_{(F3, Road)}$ 判断语义邻近的要素,本例中要素 F3 与 Road 的语义更邻近;③ 用与狭长要素语义邻近的要素缓冲区切割其他要素的缓冲区,即用 $Buffer_{(F3)}$ 切割 $Buffer_{(F2)}$,如图 5c 所示,此时缓冲区就没有重叠了。当 F2 与 F3 同类时候,用面积较大的切割面积较小的缓冲区;④ 用处理过后的缓冲区对 Road (Right) 进行分割,如图 5d 所示 Road (Right) 被分割为 Road1 与 Road2 两部

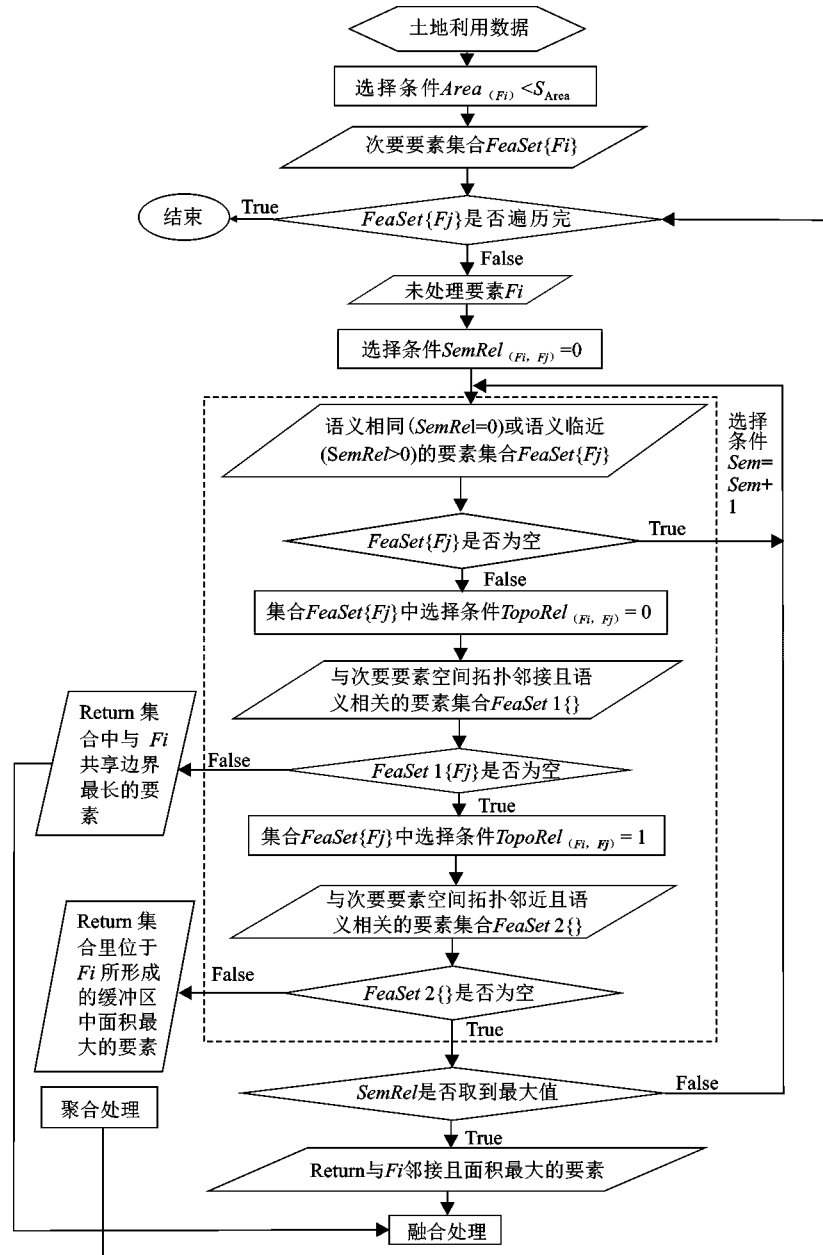


图3 模型流程图

Fig.3 Workflow of the model

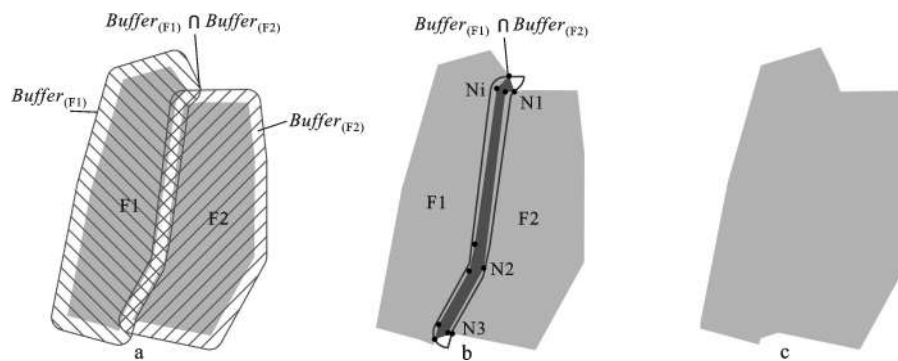


图4 图斑聚合过程

Fig.4 Process of land use patch aggregation

分;⑤ 将 Road1 与 Road2 分别合并到相应的要素中, Road1 合并入 F2, Road2 合并入 F3 中, 最终处理结果如图 5e 所示。在土地利用综合过程中, 狭长地物要经过降维处理, 将长条面要素压缩为半依比例尺的线要素, 此例中可用长条地物提取的中心线作为其线要素, 此线要素没有贯穿图斑要素, 能更好的表达该要素的拓扑位置。

1.2.3 边线化简算法

线要素化简算法中比较经典的有 Douglas-Peucker 算法^[22]、渐进式化简算法^[23]、斜拉式弯曲划分算法^[24]、Li-Open-Shaw 算法等, 本文采用 Douglas-Peucker 算法, 该算法常用于全局的线化简, 不但能保持矢量线条的形状特性, 而且可以根

据制图要求确定化简容差, 能有效的去除对要素整体形状影响微小的节点。该算法原理是首先将线条的首尾 2 端点连成一条直线, 度量 2 个端点间的每个节点到该直线的垂直距离, 若最大距离在指定容限内, 则删除首尾 2 端点间的所有节点, 若存在节点到直线的距离大于容限值, 则从该点到这条直线的首尾两端点再分别连成 2 条直线, 再分别进行比较, 直到线段不能再被划分为止(图 6)。

在对土地利用数据进行边线化简时要注意: 行政区边界、道路、河流等地理位置重要的线要进行一致性化简, 不可仅仅化简一部分, 这样会造成行政区边界不一致或者河流、道路等要素与周围要素的拓扑关系发生变化等错误。

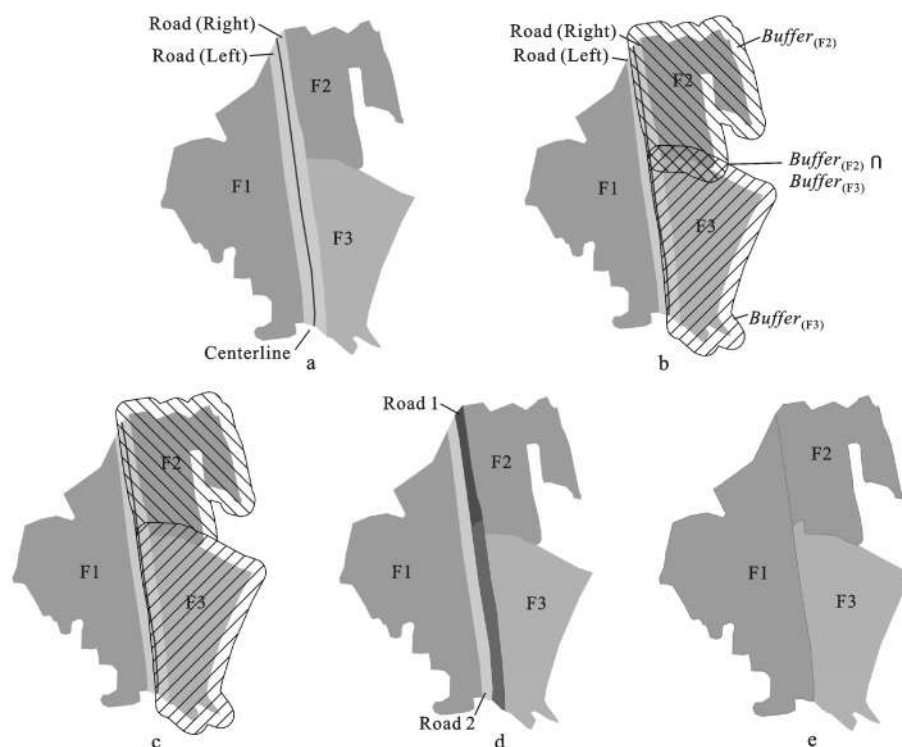


图 5 狭长要素处理过程

Fig.5 Process of handling strip feature

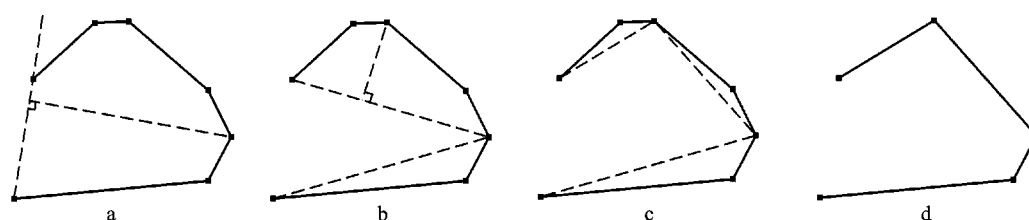


图 6 Douglas-Peucker 算法过程

Fig.6 Douglas - Peucker algorithm procedure

2 实例分析

采用辽宁省大连市旅顺口区北海街道第二次全国土地利用调查数据作为此次实验的源数据,通过规划基数转换将地类统一到中华人民共和国土地管理行业标准 TD/T 1027-2010 文件中附录 B 的土地规划地类中(图 7a),该数据最小面积为 400 m^2 ,比例尺为 1:1 万。按照 2006~2020 年土地利用总体规划 1:10 万土地利用数据的要求,采用最小上图面积为 $10\,000 \text{ m}^2$,图斑聚合最大距离 50 m 作为综合步长,运用上述综合方法进行综合,处理结果数据如图 7b。综合完成后图斑数由 1 250 个减少到 580 个,压缩比率为 53.6%,各地类面积变化如表 1 所示,可以发现综合前后除牧草地以外各地类面积变化率均小于 5%,其中较为关键的地类如城乡建设用地和耕地变化率不超过 1%,由于北海街道土地利用数据中牧草地类型的图斑只有一块且小于最小上图要求,使得周围图斑合并到牧草地中,因此综合后面积变化比例较大,属于特殊情况。地类分布及信息保留方面,相邻接的同类地块合并不会造成信息损失;相离且聚集在一起的同类碎小地块由于采用聚合处理将合并为一个大图斑,信息得到保留且位置信息准确,面积信息丢失情况与聚合距离的大小有关,聚合距离大,信息丢失多;狭长地物化简为线,面积保存在线状

地物中,因此其面积信息及位置信息保留,形状信息会损失一部分。因此综合时对地理信息的影响主要是由于离群的小斑块被合并到其他地类引起的。

表 1 综合前后地类面积变化

地类(二级)	1:10000 面积(m^2)	1:100000 面积(m^2)	变化率
城乡建设用地	3591616.8	3581008.5	0.3%
耕地	7310081.2	7263653.0	0.6%
交通水利用地	344421.8	331129.5	3.9%
林地	10688829.7	10975084.1	-2.7%
牧草地	4378.3	10685.9	-144.1%
其他建设用地	41656.6	41656.6	0.0%
其他农用地	1478613.1	1405879.1	4.9%
水域	857290.4	866583.7	-1.1%
园地	2170817.9	2137894.1	1.5%
自然保留地	3091709.5	2965840.8	4.1%

3 结 论

在计算机与 GIS 技术快速发展的今天,依赖主观意识的传统手工操作不断向客观自动处理转变,越来越多的学者投身于自动综合算法研究,新的算法层出不穷,但是对于不同类型的数据应考虑其自身的特点,选择合适的因素作为主导,并挑

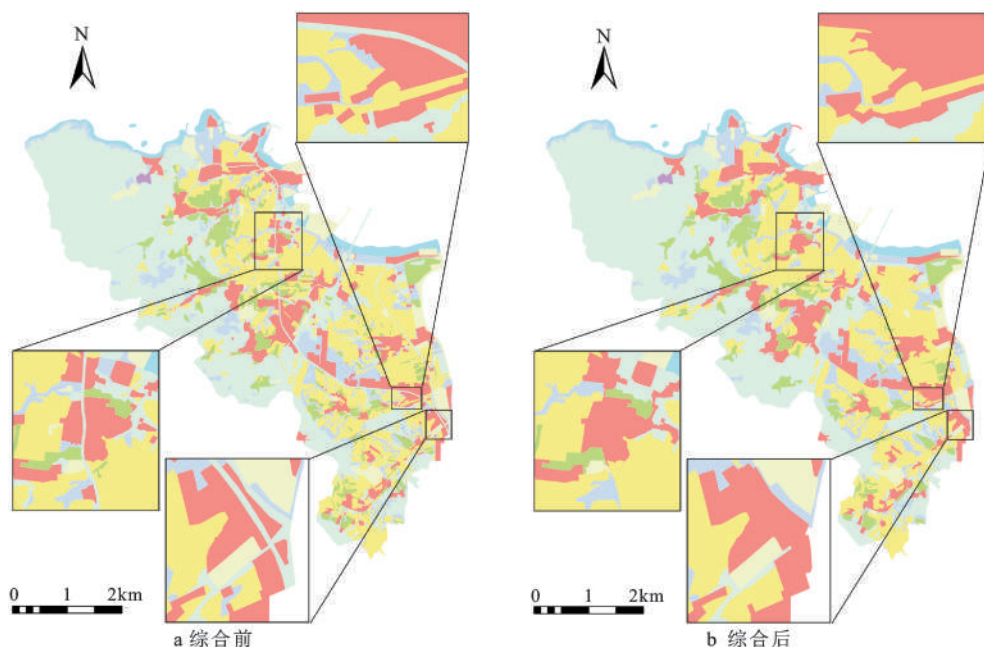


图 7 北海街道数据综合前后对比

Fig.7 The generalization result of Beihai Sub-district data

选合适的综合算法。

土地利用数据具有多语义性和空间关系复杂性,且语义起主导作用,因此土地利用数据综合应首先考虑要素语义。采用语义邻近为主导的综合方法处理土地利用数据能最大限度的保持原始数据要素分布的总体特征且能使得各个地类的总面积变化最小;另外文中采用缓冲区算法处理要素聚合与狭长地物剖分,算法简单,处理速度快,能使各类要素得到有效合理的处理。最后利用辽宁省大连市旅顺口区北海街道数据进行综合实验并分析结果,利用上述方法能有效维持各地类面积,满足综合要求。在实际实验过程中由于采用缓冲区分割狭长地物会造成分割不平滑(如图5e,要素F3吸收的狭长地物与F2接触的地方有一块小凸起区域);此外狭长要素的选择也需手动参与,这将是作者今后研究的重点。

参考文献:

- [1] 石志宽,汤国安.土地利用现状数据缩编的质量评价方法[J].地理研究,2011,30(2): 233~242.
- [2] Zhu Y, Zang D, Zhou S. Analysis on curve fit algorithm about line generalization in map generalization[C]. Chengdu, China: IEEE Computer Society, 2009.
- [3] Li W, Qiu J, Long Y. Application of rough set theory in map generalization[J]. Cehui Xuebao/Acta Geodaetica et Cartographica Sinica, 2012, 41(2): 298-301.
- [4] Dilo A, van Oosterom P, Hofman A. Constrained tGAP for generalization between scales: The case of Dutch topographic data [J]. Computers, Environment and Urban Systems, 2009, 33(5): 388-402.
- [5] 艾廷华,郭仁忠,陈晓东. Delaunay三角网支持下的多边形化简与合并[J]. 中国图象图形学报, 2001, 6(7): 93~99.
- [6] Zhu Y, Zhou S, Lu T. Research on spatial data line generalization algorithm in map generalization[J]. Journal of Software, 2011, 6(2): 241-248.
- [7] Li W, Long Y, Lin Z. The research of object selection method based on rough set in map generalization[C]. Nanjing, China: SPIE, 2007.
- [8] Harrie L. Weight-setting and quality assessment in simultaneous graphic generalization[J]. Cartographic Journal, 2003, 40(3): 221-233.
- [9] 艾廷华,毋河海. 相邻多边形共享边界的一致化改正[J]. 武汉测绘科技大学学报. 2000, 25(5): 426~431.
- [10] 刘 鹏,翁 杰,郭庆胜. 一种改进的图斑合并算法[J]. 武汉大学学报(信息科学版), 2012, 37(9): 1116~1119.
- [11] 于宗波. 试论影像地图的制图综合和编制工艺[J]. 地理科学, 1988, 8(1): 87~93.
- [12] 黄万里,戴文远,余 珊. 基于面积保持的 Douglas-Peucker 改进算法的多边形化简[J]. 科学技术与工程, 2009, 9(24): 7325~7328.
- [13] 李建林,朱德海,宋晓眉,等. 一种基于面积平衡约束的图斑化简算法[J]. 地理与地理信息科学, 2009, 25(1): 103~106.
- [14] 齐清文,刘 岳. GIS 环境下面向地理特征的制图概括的理论和方法[J]. 地理学报, 1998, 53(4): 17~27.
- [15] 艾廷华,杨 帆,李精忠. 第二次土地资源调查数据建库中的土地利用图综合缩编[J]. 武汉大学学报(信息科学版), 2010, 35(8): 887~891.
- [16] 刘耀林,李红梅,杨淳惠. 基于本体的土地利用数据综合研究 [J]. 武汉大学学报(信息科学版), 2010, 35(8): 883~886.
- [17] 艾廷华,刘耀林. 土地利用数据综合中的聚合与融合[J]. 武汉大学学报(信息科学版), 2002, 27(5): 486~492.
- [18] Qiao Q, Zhang T. Automated map generalization in distributed environments[C]. Sanya, Hainan, China: IEEE Computer Society, 2009.
- [19] Foerster T, Lehto L, Sarjakoski T, et al. Map generalization and schema transformation of geospatial data combined in a Web Service context[J]. Computers, Environment and Urban Systems, 2010, 34(1): 79-88.
- [20] Bimonte S, Bertolotto M, Gensel J, et al. Spatial OLAP and map generalization: Model and algebra[J]. International Journal of Data Warehousing and Mining, 2012, 8(1): 24-51.
- [21] 齐清文,姜莉莉. 面向地理特征的制图综合指标体系和知识法则的建立与应用研究[J]. 地理科学进展, 2001, 20(S1): 1~13.
- [22] 刘晓红,李树军,黄文骞. 制图综合中偏角限制道格拉斯算法研究[J]. 测绘与空间地理信息, 2006, 29(1): 59~60.
- [23] 郭庆胜. 线状要素图形综合的渐进方法研究[J]. 武汉测绘科技大学学报, 1998, 23(1): 54~58.
- [24] 钱海忠,武 芳,陈 波,等. 采用斜拉式弯曲划分的曲线化简方法[J]. 测绘学报, 2007, 36(4): 443~449.

Generalization of Land Use Patch Based on Semantic Priority —A Case of Beihai Sub-district of Lushun Port

YANG Jun^{1,2}, XI Jian-chao¹, KONG Fan-qiang², GE Quan-sheng¹, LI Xue-ming², GAO Hang²

(1. *Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;*

2. *Liaoning Key Laboratory of Physical Geography and Geomatics, Dalian, Liaoning 116029, China*)

Abstract: Generalization of land use patch is the key technology in the multi-scale representation of land-use data. It consists essentially of the combination of adjacent patch, the aggregation of separated semantic similar patches, the decomposition and merging of strip patches and the simplification of boundaries of features, etc. This article has gain it from the following four methods: 1) the proximity analysis model was built based on the semantic similarity of features and the spatial topological relationship. and then the nearest adjacent patch is retrieved through it to complete generalization; 2) Polygon elements are established to fill bridge areas using the nodes that contained separated factor buffer intersections; 3) The narrow regions will be split by its adjacent features buffer, and then merged with the surrounding features to integrate the regions; 4) The Douglas-Peucker algorithm is used to conduct line simplification. The above methods put respective synthesis algorithm for general and particular patches. And the experiment result from the Beihai sub-district of Lushun Port, Dalian City, shows that the models and algorithms can maximally balance area of various types of land, effectively control elements distortion, reasonably compress data to ensure the appearance of map and easiness of automatic cartography generalization.

Key words: semantic similarity; cartographic generalization; boundary simplification; land use patch