引文格式:王勇,罗安,任福,等.志愿者矢量地图的比例尺多准则估算方法[J].武汉大学学报(信息科学版),2026,51(4):815-825.DOI:10.13203/j.whugis20240342
Citation:WANG Yong,LUO An,REN Fu,et al.Multi-criteria Estimation Method for Scale of Volunteered Vector Maps[J].Geomatics and Information Science of Wuhan University,2026,51(4):815-825.DOI:10.13203/j.whugis20240342

志愿者矢量地图的比例尺多准则估算方法

王 勇[1,2,3] 罗 安[1] 任 福[2] 钟其洋[2] 廖一霖[1,3]
黄中宇[2] 郭庆胜[2]

1 中国测绘科学研究院,北京,100036
2 武汉大学资源与环境科学学院,湖北 武汉,430079
3 西南交通大学地球科学与工程学院,四川 成都,610031

摘 要:比例尺在志愿者矢量地图数据被用于多源空间数据融合时是非常重要的信息,因为比例尺差异大的空间数据在融合时易产生不一致性,会严重影响地图数据的融合质量及应用效果。因此,提出了一种基于三角形交互的消去与选择转法(elimination and choice translating reality with triangle interaction, Electre-tri)多准则决策模型的矢量地图比例尺自动估算方法。首先,分析了地理空间数据比例尺的评估参数,建立了8个空间尺度评估参数与矢量地图比例尺之间的映射关系;然后,基于Electre-tri多准则决策模型,根据偏好阈值、无差别阈值和否决阈值3类阈值动态设置,实现了待评估矢量地图与参考样本地图数据的关联计算;最后,通过不同策略准则方法实现了志愿者矢量地图的比例尺估算。为验证所提方法的有效性,以中国北京市开放街道地图的矢量地图为例,选择10个边界比例尺进行估算分析,结果表明,所提方法在志愿者矢量地图比例尺估算方面的实用性较强,模型参数的可操作性也强。

关键词:志愿者地理信息;空间尺度参数;多准则决策模型;比例尺估算
中图分类号:P208 **文献标识码**:A **收稿日期**:2025-03-19
DOI:10.13203/j.whugis20240342 **文章编号**:1671-8860(2026)04-0815-11

Multi-criteria Estimation Method for Scale of Volunteered Vector Maps

WANG Yong[1,2,3] *LUO An*[1] *REN Fu*[2] *ZHONG Qiyang*[2] *LIAO Yilin*[1,3]
HUANG Zhongyu[2] *GUO Qingsheng*[2]

1 Chinese Academy of Surveying and Mapping, Beijing 100036, China
2 School of Resources and Environmental Sciences, Wuhan University, Wuhan 430079, China
3 Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 610031, China

Abstract: **Objectives**: The data scale is one of the key factors to be considered when integrating multi-source spatial data. Especially, when volunteered vector map data is involved in the multi-source spatial data integration, due to the lack of data scale information, it may lead to data inconsistency issues. This seriously affects the quality and applicability of the fused map data. How to automatically calculate the scale of the volunteer vector data is the key to improving the quality and application value of the fusion of volunteer vector data. **Methods**: An automatic scale estimation method for vector maps based on the elimination and choice translating reality with triangle interaction (Electre-tri) multi-criteria decision model is proposed. This method mainly consists of three parts. First, by analyzing the scale evaluation parameters of geographic spatial data, we select eight typical spatial scale evaluation parameters, which include the shortest straight line segment length, the minimum bending area, the minimum size, the median side length, the vertex density, the number of spatial targets, the type of map elements, and the data capture source. And we also establish the mapping relationship between these evaluation parameters and the scale of vector maps. Then,

基金项目:国家重点研发计划(2024YFC3015603)。
第一作者:王勇,博士,研究员,主要从事地理时空大数据融合与治理等研究。wangyong@casm.ac.cn
通信作者:罗安,博士,副研究员。luoan@casm.ac.cn

we develop an Electre-tri multi-criteria decision-making model by combing the characteristics of the volunteer vector maps. Based on the decision model, we achieve the correlation calculation between the vector map data to be evaluated and the reference sample map data by dynamically setting three types of thresholds that are preference threshold, indifference threshold, and rejection threshold. Finally, by leveraging the correlation with the reference sample map data, the scale of the volunteer vector map can be estimated according to different strategy criteria methods. **Results**: To verify the effectiveness of the proposed method, we select ten types of boundary scales of OpenStreetMap data in Beijing city, China as an example. The results indicate that the proposed method has strong practicability in estimating the scale of volunteer vector maps, with high operability of model parameters. **Conclusions**: The proposed method provides an operable solution for automated scale estimation in volunteered vector maps, effectively supporting multi-source spatial data fusion by reducing scale-induced inconsistencies.

Key words: volunteered geographic information; spatial scale parameters; multi-criteria decision-making model; scale estimation

比例尺是代表矢量地图数据精细程度的重要特征,是影响多源矢量地图数据融合质量及分析效果的重要因素,对缺乏比例尺信息的矢量地图数据进行比例尺估算具有重要意义。在多源异构矢量地图数据融合和同化过程中,当不同来源的矢量地图比例尺差异大时容易产生空间数据的不一致性[1]。志愿者地理信息(volunteered geographic information, VGI)是多源地理空间数据的重要来源之一[2],是一种由非专业志愿者通过互联网平台采集、标注和共享的地理信息数据。目前广泛使用的开放街道地图(OpenStreetMap, OSM)就是VGI的典型代表[3],是一种以矢量模式存储和可视化的志愿者矢量地图。

在地理信息科学中,尺度是一个非常重要的概念,学者们从尺度的定义[4]、分类[5]、层次[6]和表达形式[7]等方面详细研究了尺度的理论问题。空间尺度是一个抽象的概念,可以用空间粒度或比例尺等指标进行描述,栅格数据的空间粒度常用分辨率进行描述,矢量数据的空间粒度则可用不同的指标如最短直线段长度、最小面积等来描述。空间粒度和比例尺之间有一种对应关系,例如,在航测地图生产中,某比例尺的地形图生产就需要选择合适的航片分辨率。当1∶5万比例尺的纸质地形图按照规范数字化为矢量空间数据时,所得矢量地图的比例尺相当于1∶5万,这个规范里就定义了相应的矢量数据的空间粒度。比例尺和尺度的英文是同一个词,但中文的比例尺习惯上就是地图比例尺。因此,矢量地图的比例尺可理解为该数据在可视化过程中的适宜比例尺范围,可以通过不同的空间粒度参数来表达。空间数据的比例尺是空间尺度必不可少的一种表达形式[8],可以用不同的参数间接描述,如空间粒度[9]、视距[10]等。

用不同类型空间粒度与矢量地图的比例尺建立函数关系是学者们常用的研究方法,例如:(1)将曲线的叶子弯曲宽度和深度的最小值作为曲线的细节层次参数,可以建立最小可分辨距离和细节层次参数与比例尺之间的关系[11];(2)不同地图目标的信息量变化与关键比例尺之间存在关系[12];(3)将方根模型与地理要素的频域信息量相结合,可以建立曲线对应比例尺的信息方根模型[13];(4)把曲线视觉粘连程度定义为视觉清晰度,可以建立多比例尺的清晰度拟合函数[14];(5)通过大批量样本统计方式可以建立河流要素比例尺变化与豪斯多夫距离中位数之间的普适函数[15];(6)以相同区域的空间数据样本库为参考,根据空间目标几何特征值变化与空间数据比例尺之间的函数关系可以推算这个区域新空间数据集的比例尺[16-17]。以上方法从不同的角度研究了某种参数指标与矢量地图比例尺之间的关系,但矢量地图数据集合可能同时包含多种类型的空间目标,不同类型的空间目标往往需要用不同的参数与比例尺建立函数关系。若空间目标的不同几何特征参数指标都与比例尺之间有函数关系,甚至同时需要考虑语义粒度与比例尺之间的关系,那么如何综合考虑这些参数对比例尺估算的影响仍是一个需要探讨的问题[18-19]。

综上所述,亟需构建一个能够综合考量多项指标的比例尺推算模型。多准则决策模型三角形交互的消去与选择转法(elimination and choice translating reality with triangle interaction, Electre-tri)是一种多准则决策排序方法。该模型能用于处理具有多个评估标准的决策问题,这些评估标准具有不确定性和非补偿性,但此模型的计算

比较复杂,参数设置较多。当然,地图比例尺从大到小也可以进行排序[19-20]。志愿者矢量地图的比例尺估算涉及多个空间尺度参数,用每一个空间尺度参数都可以估算出一个比例尺或一个比例尺范围,用不同空间尺度参数估算出的比例尺可能不同,不同空间尺度参数与比例尺之间的函数关系就是不同的评价准则。因此,矢量地图的比例尺估算问题就可以转换为一个多准则决策问题[19]。Electre多准则决策数学模型最早是由法国学者Roy于1991年提出的,研究者们在后续研究和应用中进行了改进并提出了Electre-tri模型[20],其基本思想是:首先,在若干个类别的分界点处设置边界,计算各个边界在不同评价准则上的评价得分即边界值,形成多维边界值;然后,将待分类的对象在不同评价准则上的评价得分与这些多维边界值进行比较,在综合考虑所有评价准则的评估结果后,得到待分类对象与每个边界的优劣关系;最后,依据这些优劣关系综合估算待分类对象的类别。

本文提出了一种基于Electre-tri多准则决策模型的志愿者矢量地图比例尺范围自动估算方法,先分析志愿者矢量地图的空间尺度参数,建立空间尺度参数与矢量地图比例尺之间的映射关系,再根据常用地图比例尺的矢量地图样本库来实现待评估矢量地图的比例尺范围估算。

1　空间尺度参数及其与比例尺的关系

1.1　空间尺度参数分析

地图要素的几何特征变化与地图比例尺的变化之间存在函数关系[21],在地图编制中,比例尺与地图要素基本图形认知单元的大小有关,这些基本图形认知单元主要是直线段、弯曲和简单多边形等,它们的几何特征值表达了不同方面的空间粒度,能反映地图比例尺的大小。除了空间粒度能反映地图比例尺之外,矢量地图的比例尺还可以用语义粒度、信息量和几何精度等衡量[19]。为了精准推算矢量地图的比例尺,本文选择了8个空间尺度参数,分别是最短直线段长、最小弯曲面积、最小尺寸、中位边长、顶点密度、空间目标数量、地图要素类型和数据捕获源[19]。这些空间尺度参数的含义和计算方法如下:

1)最短直线段长

矢量地图中的要素是通过一系列有序顶点构成的折线来表示,相邻顶点之间的最小距离须大于或等于地图视觉最小可分辨的长度。这种顶点之间最小的距离线段就是数字地图上最短的直线段,其长度是数字地图上线状地图要素的一个空间粒度指标,可以作为线状要素的一个空间尺度参数。

2)最小弯曲面积

在地图编制中,线弯曲是描述线状要素制图综合指标的一个重要概念[22],线弯曲的最小面积可以反映比例尺的大小。因此,最小弯曲面积是数字地图上线状要素的一个空间粒度指标,把最小弯曲面积作为线状要素矢量数据的一个空间尺度参数是合理的。线弯曲的划分方法有多种,如基于曲线轴线的划分方法[23]、基于斜拉式的弯曲划分方法[24]、基于约束Delaunay三角网的弯曲划分方法[25]、基于单调链的弯曲划分方法[26]、基于极值点的弯曲划分方法[27]等。本文选用了基于极值点的弯曲划分方法[27],因为这样划分的线弯曲与地图编制规范里描述的弯曲图形基本一致,符合地图认知规律。

3)最小尺寸

最小面积是地图面状要素综合的一个常用参数,数字地图面状目标的最小尺寸是一个最小简单面状目标的面积。最小尺寸越大,说明地图可视化表达越详细,比例尺越大。因此,最小尺寸是数字地图面状要素的一个空间粒度指标,把最小尺寸作为面状要素的一个空间尺度参数是合理的。

4)中位边长

统计整个矢量地图数据库的所有直线段的长度,将这些长度值从小到大排序,处于排序中间位置的直线段长度为中位边长。在统计规律上,中位边长是整个直线段长度数列的一个代表,能反映该矢量地图数据库生产者在曲线上采点的代表性间距,在一定程度上说明了线弯曲程度,隐含了比例尺信息。因此,中位边长是数字地图上线状要素的一个空间粒度指标,是线状要素的一个空间尺度参数。

5)顶点密度

在点状地图要素中,顶点密度是地理空间数据集中单位面积的点数;在线状地图要素和面状要素中,顶点密度是折线上单位长度的顶点数;整个矢量空间数据库的顶点密度是单位面积的所有顶点数和点状目标数的总和。在同一个地理区域中,若某类矢量地图数据的顶点密度越大,则其比例尺越大。顶点密度能反映该矢量地图数据库生产者建库时的采点密度,把顶点密度

作为矢量地图的一个空间尺度参数是合理的。

6)空间目标数量

空间目标数量是空间数据集中点状、线状和面状地图要素空间目标数量的总和。地理空间目标数量能在一定程度上反映矢量地图数据的比例尺,例如,在同一个地理区域中,不同比例尺地形图上的河流目标数量不同。因此,把空间目标数量作为矢量地图的一个空间尺度参数是合理的。

7)地图要素类型

地图要素类型能够反映矢量地图的语义粒度,可以用语义层次树结构的节点进行层次描述[3],例如,若矢量空间数据中道路的等级越低,则其比例尺越大。因此,地图要素类型能粗略反映矢量地图的比例尺范围,可以作为一个空间尺度参数。

8)数据捕获源

数据捕获源是数据的获取来源,反映了矢量地图的几何精度,可以粗略反映比例尺的大小[19],例如,全球定位系统(global positioning system, GPS)数据通常比陆地卫星(Landsat)数据集的定位精度高。因此,本文也把数据捕获源作为一个空间尺度参数。

1.2 空间尺度参数与比例尺之间的关系

本文将最小弯曲面积看作是一个半圆的面积,取该半圆的直径作为替代性参数;将最小尺寸看作是一个正方形的面积,用该正方形的边长作为替代性参数;另外,分别用顶点密度和空间目标数量的倒数作为替代性参数,参数值越大,则比例尺越小。

中国公开版的国家基础地理信息数据库中有3种比例尺,分别是1∶25万、1∶100万和1∶400万,本文用这些数据在同一个地理区域建立了3种不同比例尺大小的样本库,挖掘该区域中矢量地图数据空间尺度参数与相应样本库比例尺之间的关系。以最短直线段长为例,首先,分别计算3个比例尺样本库中线状地图要素的最短直线段长D_1、D_2和D_3,其对应的样本库比例尺分母分别为B_1、B_2和B_3;然后,定义一个比例尺分母$B_{\min}=1$对应的最短直线段长$D_{\min}=0.01$ m(实地长度);最后,用这4个样本数据建立最短直线段长与比例尺分母之间的拟合函数,可以选择幂函数或指数函数作为拟合函数,选择拟合时决定系数R^2最大的拟合函数,图1是用于拟合线性空间尺度参数与比例尺分母之间关系的幂函数。

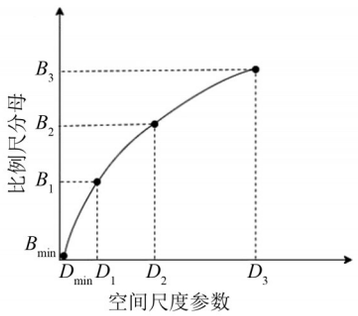


图1 线性空间尺度参数与比例尺分母之间的关系

Fig. 1 Relationship Between Linear Spatial Scale Parameters and Scale Denominators

地图要素类型是一种语义粒度参数,本文依据国家基本比例尺地图图式及其建库规范[28-29],采用等级序列值表达地图要素分类的不同抽象层次,并建立地图要素类型的等级序列值与比例尺之间的对应关系:比例尺为1∶400万的地图要素类型的等级序列值为4,比例尺为1∶25万、1∶50万和1∶100万的地图要素类型的等级序列值均为3,比例尺为1∶2.5万、1∶5万和1∶10万的地图要素类型的等级序列值均为2,比例尺为1∶5 000和1∶1万的地图要素类型的等级序列值均为1,比例尺为1∶1的地图要素类型的等级序列值为0.5。按照这种定义就可以依据矢量地图的地图要素类型所属的比例尺范围确定其等级序列值。

依据测绘生产经验,数据捕获源可以粗略地反映矢量地图的比例尺,不同来源的空间数据有不同的空间分辨率或定位精度标准,本文用等级序列值定量化不同的数据捕获源,所选的代表性数据捕获源包括公开版影像数据(如Landsat数据)、公开版国家基础地理信息数据、VGI(如OSM数据)和实测数据(如轨迹数据、GPS数据),它们的等级序列值分别为4、3、2和1,将其粗略对应的比例尺分别设置为≤1∶100万、≤1∶10万且>1∶100万、≤1∶5 000且>1∶10万、≤1∶1且>1∶5 000。

2 基于Electre-tri模型的比例尺范围估算

利用Electre-tri模型估算矢量地图比例尺范围的步骤如下:

1)定义评价准则、类别和边界

依据前文描述的空间尺度参数,本文选择了8个评价准则,分别是顶点密度准则、中位边长准则、最小弯曲面积准则、最短直线段长准则、最小

尺寸准则、空间目标数量准则、地图要素类型准则和数据捕获源准则。但是,具体的评价准则需依据矢量地图数据的几何类型进一步选取,例如,点状地图要素的评价准则就不能选择最短直线段长准则、最小弯曲面积准则和中位边长准则等,只能选择顶点密度准则、空间目标数量准则和地图要素类型准则等。

本文把比例尺的大小看成优劣关系,并按照比例尺从大到小设置类别,例如,>1:1、≤1:1且>1:5 000、≤1:5 000且>1:10万等,当然,这些类别可以进一步细分。每一个准则的相邻类别之间都存在边界,即边界比例尺或相邻类别之间的比例尺所对应的空间尺度参数值。所有准则中每个类别的边界比例尺是一样的。

一个待估算比例尺的矢量地图数据集可以依据用户需要被分解为多个子集,例如,点状地图要素子集、线状地图要素子集等。设有m个待估算比例尺的矢量地图数据子集组成的集合$A=\{a_1,a_2,\cdots,a_m\}$,每个子集就是Electre-tri模型中的一个待评价单元;n个评价准则的集合$G=\{g_1,g_2,\cdots,g_n\}$,$g_j(a_i)$表示待评价单元a_i在评价准则g_j上的评价得分;比例尺从大到小的有序类别集合$K=\{K_1,K_2,\cdots,K_q\}$,这些类别对应的边界集合$B=\{b_1,b_2,\cdots,b_h\}$,其中$h=q-1$,b_h表示两个相邻类别之间的边界。$g_j(b_h)$表示边界b_h在评价准则g_j上的评价得分,即b_h的边界值。边界集合B中每个边界有一个对应的比例尺,在评价准则g_j上也有一个边界值。类别集合K与边界集合B有一对一的映射关系,边界b_h是类别K_h的上边界,是类别K_{h+1}的下边界。

2)计算每个评价准则的边界值

比例尺本身不能作为类别的边界值,需要计算该比例尺在每个准则上对应的边界值,其计算方法如下:(1)在与待估算的矢量地图数据集相同的地理区域内,建立至少两个不同比例尺的地理空间数据样本库,本文使用的3个样本库的比例尺分别是1:25万、1:100万和1:400万。(2)依据前文所述的空间尺度参数计算方法,为每条评价准则计算每个样本库所涉及的空间尺度参数值。(3)将这些空间尺度参数值转换为评价准则中的边界值。以本文使用的3个样本库为例,用式$g_j(b_h)=f/v$计算空间目标数量准则和顶点密度准则的3个边界值,其中,v是空间目标数量或顶点密度,f是一个调整系数,本文设置$f=1\times 10^t$,t为正整数,其作用是保证这3个边界值的最小值大于1,便于与其他准则相比较。(4)按照地图要素类型等级序列值和数据捕获源等级序列值的计算方法,计算这两个评价准则上的所有边界值。(5)除地图要素类型和数据捕获源准则之外,依据已经计算出的3个样本库比例尺所对应边界的边界值,用类似图1的拟合函数插值生成所有准则的其他边界值。

3)设置无差别阈值、偏好阈值和否决阈值

为了评估待评价单元a_i的比例尺与b_h边界比例尺之间的优劣关系(即比例尺大小),需要设置无差别阈值、偏好阈值和否决阈值,它们都是待评价单元的评价得分与边界值之间的差值。为了简化这些阈值的设置,本文用边界值的百分比值表示这些阈值。

假设边界b_h在评价准则g_j上的无差别阈值为$q_j(b_h)$,偏好阈值为$p_j(b_h)$,并且$0\leqslant q_j(b_h)<p_j(b_h)$。若$g_j(a_i)-g_j(b_h)\leqslant q_j(b_h)$,则说明在第$j$个评价准则上待评价单元$a_i$的比例尺与$b_h$边界比例尺没有显著差异;反之,就需要计算它们之间的优劣关系。若$g_j(a_i)-g_j(b_h)\geqslant p_j(b_h)$,则说明待评价单元$a_i$的比例尺比$b_h$边界比例尺相比明显偏小;反之,就需要计算它们之间的优劣关系。否决阈值用于说明在第j条评价准则上是否有否决权。设否决阈值为$v_j(b_h)$,并且$v_j(b_h)>p_j(b_h)$。若$g_j(b_h)-g_j(a_i)>v_j(b_h)$,则不支持在第$j$个评价准则上待评价单元$a_i$的类别比边界$b_h$更优,说明有否决权;反之,则支持这个优劣关系的成立。

4)判断优劣关系

待评价单元与每个评价准则的各个边界之间存在的优劣关系包括:(1)$a_i\mathrm{R}b_h$,表示待评价单元a_i至少不比边界b_h差;(2)$b_h\mathrm{R}a_i$,表示边界b_h至少不比待评价单元a_i差;(3)$a_i\mathrm{S}b_h$,表示待评价单元a_i优于边界b_h;(4)$b_h\mathrm{S}a_i$,表示边界b_h优于待评价单元a_i;(5)$a_i\mathrm{I}b_h$,表示待评价单元a_i和边界b_h近似相同;(6)$a_i\mathrm{N}b_h$,表示待评价单元a_i和边界b_h无法比较。其中,最基本的优劣关系是$a_i\mathrm{R}b_h$和$b_h\mathrm{R}a_i$。由于每个评价准则都可以推算出一个优劣关系,因此该模型用一个可信度指标反映$a_i\mathrm{R}b_h$和$b_h\mathrm{R}a_i$的总体可信程度或总体支持强度,$a_i\mathrm{R}b_h$和$b_h\mathrm{R}a_i$的可信度分别表示为$\sigma(a_i,b_h)\in[0,1]$和$\sigma(b_h,a_i)\in[0,1]$[20]。为了明确这些优劣关系是否成立,需要设置截止水平λ,通常$\lambda\in[0.5,1]$[20]。

描述$a_i\mathrm{R}b_h$成立的总体可信度$\sigma(a_i,b_h)$的计算公式为:

$$\sigma(a_i,b_h)=c(a_i,b_h)\prod_{j\in\bar{T}}\frac{1-d_j(a_i,b_h)}{1-c(a_i,b_h)} \quad (1)$$

式中,$c(a_i,b_h)$表示所有评价准则总体上支持a_iRb_h的一致性强度;$d_j(a_i,b_h)$表示待评价单元a_i和边界b_h在第j条评价准则上支持a_iRb_h的不一致性强度;$\bar{T}=\{j\in T,d_j(a_i,b_h)>c(a_i,b_h)\}$,$T$是所有评价准则;$\prod$是连乘符号。按照同样的方法可以计算$\sigma(b_h,a_i)$。

$c_j(a_i,b_h)$、$c(a_i,b_h)$和$d_j(a_i,b_h)$的计算公式分别为:

$$c_j(a_i,b_h)=\begin{cases}1, g_j(b_h)-g_j(a_i)\leqslant q_j(b_h)\\ 0, g_j(b_h)-g_j(a_i)\geqslant p_j(b_h)\\ \dfrac{p_j(b_h)-[g_j(b_h)-g_j(a_i)]}{p_j(b_h)-q_j(b_h)}, \text{其他}\end{cases} \quad (2)$$

$$c(a_i,b_h)=\sum_{j=1}^{n}w_jc_j(a_i,b_h) \quad (3)$$

$$d_j(a_i,b_h)=\begin{cases}1, g_j(b_h)-g_j(a_i)>v_j(b_h)\\ 0, g_j(b_h)-g_j(a_i)\leqslant p_j(b_h)\\ \dfrac{[g_j(b_h)-g_j(a_i)]-p_j(b_h)}{v_j(b_h)-p_j(b_h)}, \text{其他}\end{cases} \quad (4)$$

式中,$c_j(a_i,b_h)$表示待评价单元a_i和边界b_h在评价准则g_j上支持a_iRb_h的一致性强度;w_j表示第j条评价准则的权重。

5)估算待评价单元的比例尺范围

将待评价单元a_i的比例尺具体划分到哪个已经定义的类别(比例尺范围)有两种方法,分别是悲观准则方法和乐观准则方法。悲观准则方法倾向于将待评价单元a_i划分到更差的类别即更大比例尺的范围中,而乐观准则方法倾向于将待评价单元划分到更好的类别即更小比例尺的范围中。

(1)悲观准则方法:依次比较待评价单元a_i与各边界$\{b_h,b_{h-1},\cdots,b_1\}$之间的优劣关系,在每次循环中,若$\sigma(a_i,b_h)\geqslant\lambda$且$\sigma(b_h,a_i)<\lambda$,则$a_iSb_h$成立,待评价单元$a_i$的类别应该为$K_{h+1}$,并退出循环,否则执行下一个循环。若无任何边界满足条件,则需要调整λ,再进行循环判断,所推算出的类别就是待评价单元a_i的比例尺范围。

(2)乐观准则方法:依次比较待评价单元a_i与各边界$\{b_1,b_2,\cdots,b_h\}$之间的优劣关系,在每次循环中,若$\sigma(b_h,a_i)\geqslant\lambda$且$\sigma(a_i,b_h)<\lambda$,则$b_hSa_i$成立,待评价单元$a_i$的类别应该为$K_h$,并退出循环,否则执行下一个循环。若无任何边界满足条件,则需要调整λ,再进行循环判断,所推算出的类别就是待评价单元a_i的比例尺范围。

3 比例尺估算实验与分析

3.1 数据处理流程

待估算比例尺的志愿者矢量地图是从网络下载的北京市OSM矢量空间数据集,样本库是从国家基础地理信息数据库(https://www.webmap.cn/main.do method=index)中下载的公开版北京市1:25万、1:100万和1:400万3种比例尺的矢量地图数据。实验数据的处理流程如图2所示。

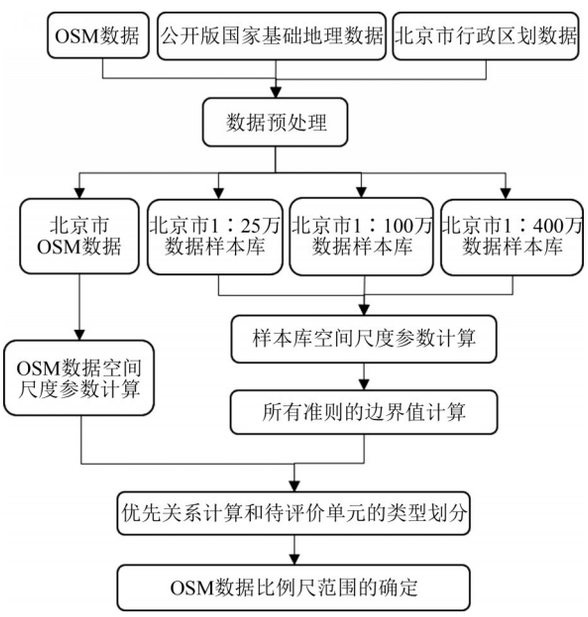
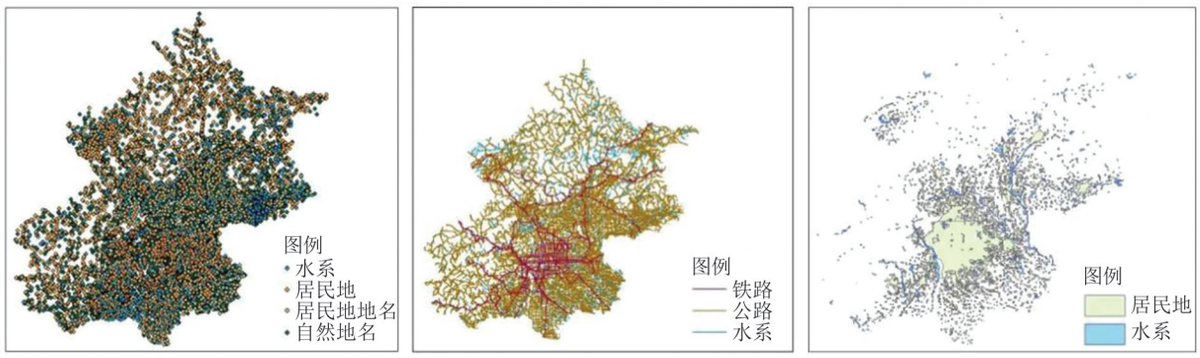


图2 实验数据处理流程

Fig. 2 Experimental Data Processing Flow

下载的实验数据中存在共线或重复的顶点等冗余点,地图投影也不一致,因此,用ArcGIS 10.2的顶点删除工具删除了共线的顶点,设置的顶点简化容差为地图上0.01 mm,用地图投影转换工具统一了几个实验数据集的地图投影。把实验数据按照点状要素、线状要素和面状要素进行分类,根据北京市行政区划边界对地图数据进行裁剪,建立北京市1:25万、1:100万和1:400万3种比例尺的点状要素、线状要素和面状要素样本数据库。例如,图3是1:25万北京市地理空间数据,图4是北京市OSM数据。本文选择这3个比例尺的地理空间数据建立样本数据库的原因是:这些数据符合地图规范且数据也可以公开。

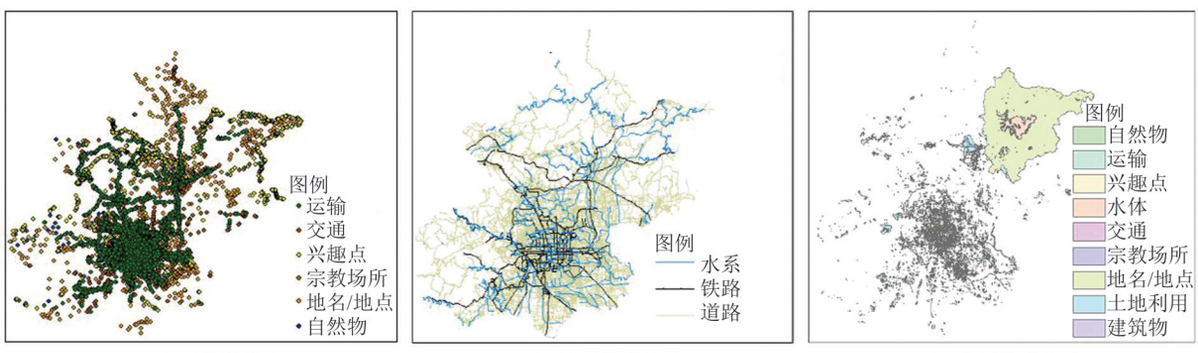
用Python语言在PyCharm平台上实现估算比例尺范围的Electre-tri模型。参考中国常用的系列地形图比例尺,设置比例尺范围分类的边界比例尺集合$B=\{1:1,1:5\,000,1:1万,1:2.5万,1:5万,1:10万,1:25万,1:50万,1:100万,1:400万\}$。



(a) 点状要素数据　　(b) 线状要素数据　　(c) 面状要素数据

图 3　1:25万北京市地理空间数据

Fig. 3　Beijing City Geospatial Data at 1:250 000



(a) 点状要素数据　　(b) 线状要素数据　　(c) 面状要素数据

图 4　北京市 OSM 数据

Fig. 4　OSM Data of Beijing City

3.2　边界值的计算

依据前文所述的矢量地图数据空间尺度参数值的计算方法,可以计算出 8 个评价准则分别在样本库 3 个比例尺边界上的边界值(即边界评价得分),具体结果见表 1。然后,可按照同样的原理,设置在比例尺 1:1 边界上的边界值。

表 1　从样本数据中提取的在所有评价准则上的边界值

Table 1　Boundary Values on All Evaluation Criteria Extracted from Sample Data

比例尺	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
1:400万	135.14	1 154.13	79.20	42.20	5 221.11	617.28	4.00	4.00
1:100万	18.52	403.16	16.83	11.15	510.83	21.00	3.00	4.00
1:25万	8.18	68.86	3.65	2.57	41.82	2.64	3.00	3.00

依据每条评价准则的 4 个已知边界值,可以为前 6 个评价准则分别建立对应的拟合函数,并推算出前 6 个评价准则中其余比例尺对应的边界评价得分,具体得分见表 2。第 7 个和第 8 个评价准则的边界得分依据前文所述方法设置。另外,在顶点密度准则和空间目标数量准则中边界值计算函数的转换系数 f 分别设为 100 和 10 000。依据空间尺度参数的计算方法,可以得到北京市 OSM 数据集待评价单元在各个评价准则上的边界评价得分,具体见表 3。

3.3　实验结果分析

在实验中设截止水平 $\lambda=0.75$,在 Electre-tri 模型中要求对每个评价准则的每个边界都设置无差别阈值和偏好阈值,但是考虑到用于边界值计算的空间尺度参数已经线性化,比例尺分母需要取整,为了简化模型中几个阈值的设置,把每个评价准则的每个边界的无差别阈值和偏好阈值进行统一设置。例如,设无差别阈值为相邻边界值差异的 5%(简称为“无差别阈值为 5%”),设偏好阈值为相邻边界值差异的 20%(简称为“偏好阈值为 20%”),设所有评价准则的权重一样,用悲观准则方法和乐观准则方法分别估算北京市 OSM 数据的比例尺范围(见表 4)。在实验中发现的规律如下:

表2 在所有评价准则上的边界评价得分

Table 2 Boundary Evaluation Scores on All Evaluation Criteria

评价准则	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
g_1	0.01	0.17	0.34	0.83	1.64	3.28	8.18	11.63	18.52	135.14
g_2	0.01	1.39	2.76	6.89	13.78	27.55	68.86	180.29	403.16	1 154.13
g_3	0.01	0.15	0.22	0.44	0.79	1.51	3.65	8.04	16.83	79.20
g_4	0.01	0.06	0.11	0.27	0.52	1.03	2.57	5.43	11.15	42.20
g_5	0.01	0.93	1.77	4.27	8.44	16.79	41.82	198.16	510.83	5 221.11
g_6	0.01	0.06	0.12	0.27	0.54	1.06	2.64	8.76	21.00	617.28
g_7	0.50	1.00	1.00	2.00	2.00	2.00	3.00	3.00	3.00	4.00
g_8	1.00	2.00	2.00	2.00	2.00	3.00	3.00	3.00	4.00	4.00

表3 北京市OSM数据在每个评价准则上的评价得分

Table 3 Evaluation Scores of Beijing OSM Data on Each Evaluation Criterion

评价准则	评价得分
g_1	0.33
g_2	41.46
g_3	0.31
g_4	0.21
g_5	3.00
g_6	0.06
g_7	1.00
g_8	2.00

1)当权重保持不变且无差别阈值为5%、否决阈值为60%时,只改变偏好阈值,估算结果不变,这说明当前数据对偏好阈值不敏感。

2)在偏好阈值为30%或50%、否决阈值为60%的条件下,只改变无差别阈值,权重不变,估算结果也没有改变,这说明当前数据对无差别阈值不敏感。

3)在无差别阈值为5%、偏好阈值为30%和否决阈值为60%的条件下,改变不同准则的权重,当顶点密度准则和空间目标数量准则的权重为35%、其他准则的权重为5%时,用悲观准则方法估算的结果与表4中的结果不同,结果为1:1万~1:5 000,但是用乐观准则方法估算的结果没有变化,这说明当前的实验数据对权重不是非常敏感。

4)在权重不变且无差别阈值为5%、偏好阈值为20%时,由表4可知,当否决阈值为50%或60%时,估算结果发生了变化,这说明否决阈值对当前的实验数据在该数值上存在突变性的影响。

因此,在北京市OSM数据的比例尺范围估算中比较合理的几种阈值设置为:无差别阈值为5%,偏好阈值为30%,否决阈值为60%。用这3个阈值设置,依据地图要素类型选择不同评价准则子集,对北京市OSM数据的全集(包含点状、线状和面状地图要素)、点状地图要素数据子集、线状地图要素数据子集和面状地图要素数据子集分别进行比例尺范围估算,估算出的比例尺范围如表5所示。这些结果说明,依据悲观准则方法估算出的这几个数据子集的比例尺范围比较一致。针对北京市OSM数据全集,从8个评价准则中选择部分评价准则对比发现(见表6),依据悲观准则方法估算的比例尺范围比较一致。

表4 基于不同否决阈值的比例尺范围估算结果

Table 4 Scale Range Estimation Results Based on Different Rejection Thresholds

否决阈值占边界值的比例/%	悲观准则方法的估算结果	乐观准则方法的估算结果
30	1:1万~1:5 000	1:25万~1:10万
40	1:1万~1:5 000	1:25万~1:10万
50	1:1万~1:5 000	1:25万~1:10万
60	1:2.5万~1:1万	1:10万~1:5万
70	1:2.5万~1:1万	1:10万~1:5万
80	1:2.5万~1:1万	1:10万~1:5万
90	1:2.5万~1:1万	1:10万~1:5万
100	1:2.5万~1:1万	1:10万~1:5万

利用文献[17]的方法,分别依据本实验3个不同比例尺样本库中空间数据的顶点密度、中位边长、最小弯曲面积、最短直线段长、最小尺寸、空间目标数量等6个空间尺度参数值x建立了6个拟合函数,分别为$f(x)=8.73\times10^4x^{0.78}$,$f(x)=481.71x^{1.28}$,$f(x)=7.94\times10^4x^{0.90}$,$f(x)=8.41\times10^4x^{1.03}$,$f(x)=2.51\times10^4x^{0.59}$,$f(x)=2.43\times10^5x^{0.44}$。用这些拟合函数计算出的北京市OSM数据的比例尺分别是1:36 767、1:56 670、1:27 672、1:16 853、1:47 993、1:70 648,对应的比例尺范围分别是1:5万~1:2.5万、1:10万~1:5万、

表5　不同矢量空间数据集的估算结果

Table 5　Estimation Results of Different Vector Spatial Datasets

待评价单元	选择的评价准则	悲观准则方法的估算结果	乐观准则方法的估算结果
北京市OSM数据	8项评价准则	1:2.5万～1:1万	1:10万～1:5万
北京市OSM点状要素数据	顶点密度评价准则	1:1万～1:5 000	1:2.5万～1:1万
北京市OSM线状要素数据	中位边长评价准则、最小弯曲面积评价准则	1:2.5万～1:1万	1:25万～1:10万
北京市OSM面状要素数据	最短直线段长评价准则、最小尺寸评价准则	1:2.5万～1:1万	1:2.5万～1:1万

表6　基于不同评价准则的北京市OSM数据子集的估算结果

Table 6　Estimation Results of Beijing OSM Data Subsets Based on Different Evaluation Criteria

未选的评价准则	悲观准则方法的估算结果	乐观准则方法的估算结果
顶点密度	1:2.5万～1:1万	1:10万～1:5万
中位边长	1:2.5万～1:1万	1:2.5万～1:1万
最小弯曲面积	1:2.5万～1:1万	1:10万～1:5万
最短直线段长	1:2.5万～1:1万	1:10万～1:5万
最小尺寸	1:2.5万～1:1万	1:10万～1:5万
空间目标数量	1:2.5万～1:1万	1:10万～1:5万
数据捕获源	1:2.5万～1:1万	1:10万～1:5万
地图要素类型	1:2.5万～1:1万	1:10万～1:5万
空间目标数量、地图要素类型、数据捕获源	1:2.5万～1:1万	1:10万～1:5万
顶点密度、中位边长、最小弯曲面积、最短直线段长、最小尺寸	1:5 000～1:1	1:1万～1:5 000
地图要素类型、数据捕获源	1:1万～1:5 000	1:25万～1:10万

1:5万～1:2.5万、1:2.5万～1:1万、1:5万～1:2.5万、1:10万～1:5万。这说明计算出的比例尺范围差异比较大,使用本文所提出的模型就可以估算北京市OSM数据的比例尺范围。

4　结　语

地理空间数据的尺度在地理信息融合和地理信息分析中是一个非常重要的概念[6-7,30],当所使用的地理空间数据没有比例尺的元数据说明时,很难保证不同比例尺空间数据之间的一致性[1]。本文引入Electre-tri多准则决策模型,通过建立空间尺度评估参数与不同策略准则方法,实现了志愿者矢量地图比例尺的自动估算,并在实验中验证了本文方法的实用性和可操作性。在本文的研究过程中认识到有些问题需要进一步探讨,主要包括:

1)空间尺度不同参数的应用。在地理分析中有时只考虑用某种空间粒度或语义粒度的空间数据,例如,变量的空间自相关程度往往依赖空间粒度与单元划分方法[30]。因此,不同空间粒度或空间尺度和不同语义粒度的VGI在空间分析中的可塑性面积单元问题是需要进一步研究的问题。

2)Electre-tri模型中比例尺范围的确定。针对不同的应用场景需要有不同的比例尺范围划分方法,例如,研究城市区域的详细规划和总体规划就需要用不同比例尺的地图。因此,本文所建立的Electre-tri模型中的比例尺系列也需要根据不同应用场景来进行修改。

3)Electre-tri模型中阈值和权重的设置。本文主要依靠经验设置Electre-tri模型中的阈值和权重,后续研究可以依据具体的地理信息应用场景来分析阈值对比例尺的影响程度,例如,可引入层次分析或权重计算方法。

参　考　文　献

[1] TOUYA G, BRANDO-ESCOBAR C. Detecting Level-of-Detail Inconsistencies in Volunteered Geographic Information Data Sets[J]. *Cartographica*, 2013, 48(2): 134-143.

[2] 张博, 张猛, 王非, 等. VGI数据与地形图数据的自动融合研究[J]. 武汉大学学报(信息科学版), 2019, 44(11): 1708-1714.
ZHANG Bo, ZHANG Meng, WANG Fei, et al. Automatic Data Integration Between VGI and Topographic Data[J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(11): 1708-1714.

[3] 杨伟, 艾廷华. 运用Delaunay三角网提取OpenStreetMap主干道多边形[J]. 武汉大学学报(信息

科学版), 2018, 43(11): 1725-1731.
YANG Wei, AI Tinghua. Extracting Arterial Road Polygon from OpenStreetMap Data Based on Delaunay Triangulation[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(11): 1725-1731.

[4] 刘凯, 毋河海, 艾廷华, 等. 地理信息尺度的三重概念及其变换[J]. 武汉大学学报(信息科学版), 2008, 33(11): 1178-1181.
LIU Kai, WU Hehai, AI Tinghua, et al. Three-Tiered Concepts of Scale of Geographical Information and Its Transformation[J]. *Geomatics and Information Science of Wuhan University*, 2008, 33(11): 1178-1181.

[5] GOODCHILD M F. Scale in GIS: An Overview[J]. *Geomorphology*, 2011, 130(1/2): 5-9.

[6] 李志林, 王继成, 谭诗腾, 等. 地理信息科学中尺度问题的30年研究现状[J]. 武汉大学学报(信息科学版), 2018, 43(12): 2233-2242.
LI Zhilin, WANG Jicheng, TAN Shiteng, et al. Scale in Geo-information Science: An Overview of Thirty-Year Development[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 2233-2242.

[7] 李志林. 地理空间数据处理的尺度理论[J]. 地理信息世界, 2005(2): 1-5.
LI Zhilin. A Theoretical Discussion on the Scale Issue in Geospatial Data Handling[J]. *Geomatics World*, 2005(2): 1-5.

[8] DABIRI Z, BLASCHKE T. Scale Matters: A Survey of the Concepts of Scale Used in Spatial Disciplines[J]. *European Journal of Remote Sensing*, 2019, 52(1): 419-434.

[9] 刘凯, 龙毅, 秦耀辰. 论地理信息的空间粒度[J]. 地理与地理信息科学, 2014, 30(1): 8-12.
LIU Kai, LONG Yi, QIN Yaochen. On the Spatial Granularity of Geographical Information[J]. *Geography and Geo-Information Science*, 2014, 30(1): 8-12.

[10] 王培, 江南, 王玉晶, 等. 利用视距进行多源空间数据尺度集成[J]. 武汉大学学报(信息科学版), 2020, 45(7): 1008-1014.
WANG Pei, JIANG Nan, WANG Yujing, et al. Spatial Scale Integration Method Based on Stadia[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(7): 1008-1014.

[11] 刘鹏程, 李精忠, 艾廷华. 数字地图要素的细节层次一致性评价模型研究[J]. 中国矿业大学学报, 2019, 48(2): 437-444.
LIU Pengcheng, LI Jingzhong, AI Tinghua. Evaluation Model of Level of Detail Consistency of the Features on Digital Maps[J]. *Journal of China University of Mining & Technology*, 2019, 48(2): 437-444.

[12] 李雯静, 毋河海. 地图信息衰减中关键比例尺的研究[J]. 武汉大学学报(信息科学版), 2006, 31(12): 1116-1119.
LI Wenjing, WU Hehai. On Catastrophe-Scale in Course of Map-Information Attenuation[J]. *Geomatics and Information Science of Wunan University*, 2006, 31(12): 1116-1119.

[13] 肖天元, 刘鹏程, 艾廷华, 等. 一种傅里叶信息度量的曲线分形描述与多尺度表达方法[J]. 武汉大学学报(信息科学版), 2020, 45(1): 119-125.
XIAO Tianyuan, LIU Pengcheng, AI Tinghua, et al. A Fractal Description and Multi-scale Expression Method of Fourier Information Metrics[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(1): 119-125.

[14] 安晓亚, 成晓强. 矢量曲线的视觉清晰度及在网络地图综合中的应用[J]. 测绘学报, 2020, 49(2): 245-255.
AN Xiaoya, CHENG Xiaoqiang. Visual Clarity of Vector Curve and Its Application in Web Map Generalization[J]. *Acta Geodaetica et Cartographica Sinica*, 2020, 49(2): 245-255.

[15] 张鑫港, 闫浩文. 河流要素比例尺变化与中位数Hausdorff距离的关系[J]. 测绘学报, 2023, 52(8): 1364-1374.
ZHANG Xin'gang, YAN Haowen. Relationship Between Scale Change and Median Hausdorff Distance of River Elements[J]. *Acta Geodaetica et Cartographica Sinica*, 2023, 52(8): 1364-1374.

[16] 樊玉鑫, 郭庆胜, 王勇, 等. 参考样本库的线状道路空间数据比例尺估算[J]. 测绘科学, 2022, 47(7): 208-215.
FAN Yuxin, GUO Qingsheng, WANG Yong, et al. Scale Estimation for Linear Road Spatial Data on Sample Database[J]. *Science of Surveying and Mapping*, 2022, 47(7): 208-215.

[17] 钟其洋, 郭庆胜, 王勇, 等. 基于空间粒度的矢量空间数据比例尺估算方法[J]. 测绘通报, 2022(8): 117-122.
ZHONG Qiyang, GUO Qingsheng, WANG Yong, et al. The Method for Estimating the Scale of Vector Spatial Data Based on Spatial Granularity[J]. *Bulletin of Surveying and Mapping*, 2022(8): 117-122.

[18] LAURINI R. A Conceptual Framework for Geographic Knowledge Engineering[J]. *Journal of Visual Languages and Computing*, 2014, 25(1): 2-19.

[19] TOUYA G, REIMER A. Inferring the Scale of OpenStreetMap Features[C]//ARSANJANI J J,

ZIPF A, MOONEY P, et al. OpenStreetMap in GIScience. Cham: Springer International Publishing, 2015: 81-99.

[20] MOUSSEAU V, FIGUEIRA J, NAUX J P. Using Assignment Examples to Infer Weights for ELECTRE TRI Method: Some Experimental Results[J]. *European Journal of Operational Research*, 2001, 130(2): 263-275.

[21] 毋河海. GIS与地图信息综合基本模型与算法[M]. 武汉: 武汉大学出版社, 2012.
WU Hehai. Basic Model and Algorithm of GIS and Map Information Integration[M]. Wuhan: Wuhan University Press, 2012.

[22] 祝国瑞, 郭礼珍, 尹贡白, 等. 地图设计与编绘[M]. 武汉: 武汉大学出版社, 2001.
ZHU Guorui, GUO Lizhen, YIN Gongbai, et al. Map Design and Compilation[M]. Wuhan: Wuhan University Press, 2001.

[23] 刘鹏程, 杨文博, 李畅. 曲线弯曲多叉树模型及其在地图综合中的应用[J]. 计算机应用研究, 2012, 29(7): 2793-2795.
LIU Pengcheng, YANG Wenbo, LI Chang. Multi-branch Tree Model of Bend for Curve and Its Application in Map Generalization[J]. *Application Research of Computers*, 2012, 29(7): 2793-2795.

[24] 钱海忠, 孙建东, 武芳, 等. 采用斜拉式弯曲划分的曲线化简方法[C]//中国测绘学会学术年会, 中国桂林, 2008.
QIAN Haizhong, SUN Jiandong, WU Fang, et al. Simplifying Line with Oblique Dividing Curve Method[C]//Academic Annual Conference of the Chinese Society for Geodesy, Photogrammetry and Cartography, Guilin, China, 2008.

[25] 艾廷华, 郭仁忠, 刘耀林. 曲线弯曲深度层次结构的二叉树表达[J]. 测绘学报, 2001, 30(4): 343-348.
AI Tinghua, GUO Renzhong, LIU Yaolin. A Binary Tree Representation of Curve Hierarchical Structure in Depth[J]. *Acta Geodaetica et Cartographica Sinica*, 2001, 30(4): 343-348.

[26] 罗广祥, 祝国瑞, 毋河海, 等. 坐标单调性分析下地图曲线弯曲识别模型的研究[J]. 测绘通报, 2005(10): 21-24.
LUO Guangxiang, ZHU Guorui, WU Hehai, et al. The Study of the Model of Identifying Bends for Cartographic Linear Features Based on Analysing Coordinate Monotony[J]. *Bulletin of Surveying and Mapping*, 2005(10): 21-24.

[27] 郭庆胜, 黄远林, 章莉萍. 曲线的弯曲识别方法研究[J]. 武汉大学学报(信息科学版), 2008, 33(6): 596-599.
GUO Qingsheng, HUANG Yuanlin, ZHANG Liping. The Method of Curve Bend Recognition[J]. *Geomatics and Information Science of Wunan University*, 2008, 33(6): 596-599.

[28] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. 基础地理信息要素分类与代码: GB/T 13923—2006[S]. 北京: 中国标准出版社, 2006.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Specifications for Feature Classification and Codes of Fundamental Geographic Information: GB/T 13923—2006[S]. Beijing: China Standards Press, 2006.

[29] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. 国家基本比例尺地图图式 第4部分:1∶250 000 1∶500 000 1∶1 000 000地形图图式: GB/T 20257.4—2017[S]. 北京: 中国标准出版社,2017.
General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Cartographic Symbols for National Fundamental Scale Maps: Part 4: Specifications for Cartographic Symbols 1∶250 000 1∶500 000 & 1∶1 000 000 Topographic Maps: GB/T 20257.4—2017[S]. Beijing: China Standards Press, 2017.

[30] 陈江平, 张瑶, 余远剑. 空间自相关的可塑性面积单元问题效应[J]. 地理学报, 2011, 66(12): 1597-1606.
CHEN Jiangping, ZHANG Yao, YU Yuanjian. Effect of MAUP in Spatial Autocorrelation[J]. *Acta Geographica Sinica*, 2011, 66(12): 1597-1606.