

顾及网络特征的复杂人工河网的自动选取

武芳 谭笑 王辉连 翟仁健

(信息工程大学测绘学院, 郑州 450052)

摘要 针对人工河网选取过程中需要考虑的几个重要因素,引入了线状要素动态分段模型组织人工河网数据及其空间知识。提出一种基于栅格的河网密度探测方法,较传统的计算网眼面积的方法简便且效果较好。建立了基于“知识决策”的人工河网选取模型,模型中考虑了复杂人工河网综合中河渠的属性(主要、次要)、长度、连通情况、河渠间的空间关系以及河网密度等5个方面的因素,建立多准则决策模型对每条备选河渠平等地进行综合评价,根据评价结果按照选取可能性从小到大间隔选取,从而能够较好保持人工河网的网络特征。

关键词 自动综合 空间关系 多准则决策 动态分段

中图分类号: P283.1 文献标识码: A 文章编号: 1006-8961(2007)06-1103-07

Study on Automated Canal Selection

WU Fang, TAN Xiao, WANG Hui-lian, ZHAI Ren-jian

(Institute of Surveying and Mapping, Information Engineering University, Zhengzhou 450052)

Abstract This paper analyzes the character of the complex canal network, and imports the dynamic segmentation theory to organize the canal network data and its spatial knowledge. A method based on grid to detect the network density is put forward and the result show it is better than other method. And also a canal network selection model is set up based on decision of spatial knowledge. This model integrates five factors of the canal selection such as the attribute, length, link condition, spatial relation between canals and network density. Then, utilizes the multiple criteria decision to evaluate each canal by these five factors, and make alternated selection to preserve the whole network character.

Keywords automated map generalization, spatial relationship, multiple criteria decision model, dynamic Segmentation

1 引言

制图综合是对地理空间对象随尺度变化过程中的删减进行合乎制图学理论要求的一种综合变换。该变换包括选取、化简、概括、位移等基本操作,其中选取是最主要的和最基本的方法,是根据制图综合指标,从大量的制图对象中选取较大或较重要的,舍去次要的非本质的地物和现象,从而使制图对象在数量上满足GIS中空间数据多尺度表达或地图生产编绘自动化的需要。在数字条件下,让计算机去自动执行这一过程是非常困难的,因为

制图综合对人脑的判断具有高度的依赖性,需要对待综合对象的空间分布格局、时空复合特征等进行结构识别;对适用的综合措施进行过程判断;对所需使用的操作模型和知识法则做出工具选择;对综合结果从地图适应性、地物间协调性、表达清晰性、内容完整性等方面进行综合评定等等。因此,自动综合的研究一直是国内外制图学界和GIS领域研究的热点。

地图上各不同要素的选取,因不同要素的图形分布特征、重要性的判定条件等的不同,在编绘规范和地图自动制图综合的过程中所采用的方法是根本不同的。在我国广袤的东部平原,分布着大量的人

基金项目:国家自然科学基金项目(40471115,40671162)

收稿日期:2005-09-06;改回日期:2006-03-15

第一作者简介:武芳(1964~),女,教授,博士生导师。研究方向为地理信息系统与辅助决策、地图自动综合等。E-mail: lzhiping@public.zz.ha.cn

工河网,这些人工河网以沟渠的形式出现,同人类的生产活动关系极为密切。图 1 是我国东部某地区 1:25 万水网图,从中可以看出人工河网的选取具有以下特点:

(1) 整体图形特征方面,人工河网呈网格状分布,河渠间往往呈直角相交,而天然河系中的河流大多呈锐角相交。天然河系的主支流关系明显,人工河网则多呈格状分布,无明显主支流关系。

(2) 就单个目标而言,人工河渠比较平直,弯曲较少,而天然河流由于受到地形的影响,弯曲较多。

(3) 人工河网只分主要、次要两个等级,而天然河系则可依据长度划分多个等级。

(4) 人工河网不能机械地按长度选取,应着重水网的整体结构特征、内部形态与网眼大小、密度对比、外围轮廓以及主、支渠内外连通情况^[1,2],而天然河系则可依据长度及其分布特征进行取舍。



图 1 复杂人工河网实验数据
Fig. 1 The complex canal network

在数字条件下,制图学者对网状地图要素的自动选取研究已比较深入,但多见于道路网的自动取舍,而对复杂人工河网自动选取国内外很少见到。因此本文根据人工河网选取的特点,提出了基于动态分段思想的人工河网数据模型,在获取人工河网空间知识的基础上,建立了人工河网的评价决策模型。本文的研究重点是如何保持人工河网选取前后的空间分布密度差异和网络特征。

2 基于动态分段的人工河网数据模型

与弧段-结点模型的静态分段不同,动态分段方

法不是在线状要素沿线某属性发生变化的地方进行物理分段,而是将属性的沿程变化存储为独立的属性表,在显示和分析中直接依据事件属性表中的距离值对线状要素进行动态逻辑分段。动态分段实质上是建立在弧段-结点数据结构上的一种抽象方法,使用弧(arc)、段(section)、路径(route)、事件(event)等实体,建立动态段与弧段-结点模型的映射关系,使用路径实体将线状要素作为整体进行描述,使用事件描述线状要素各种属性的沿程变化或者各种现象在线状要素沿线的分布^[3]。

动态分段模型首先要建立完整的路径系统,其次是建立关联到路径上的事件表。在此基础上,可进行动态分段的 3 种操作:联结、分段和分析。使用动态分段技术,可以将多组属性数据关联到一个线状特征的任何部分,可以存储、显示、查询和分析这些属性却不影响其关联的线状特征的坐标。

在人工河网选取中,主要考虑的是河渠的属性(主要、次要)、长度、连通情况、河渠间的空间关系及河网密度等 5 个方面的因素,重点是河渠间空间关系(主要有右相交、左相交、交叉和相连 4 种)和河网密度的描述。由于河渠是有一定长度的、线状延伸的,它所流经的区域可能存在着不同的密度划分,因此我们引入动态分段技术思想来组织人工河网数据模型。

图 2 是河渠 AC 流经不同密度的人工河网区域,经密度探测可知 B 点为两个不同密度区域在河渠 AC 上的分界点,因此,可以分别记录段 AB、BC 的长度及其密度等级,存储于河渠 AC 的数据结构中。在对河渠 AC 整体进行的密度评价时,用 AB、BC 的密度等级分别乘以 AB、BC 段所占整个河渠 AC 的长度比例,即可得河渠 AC 的整体密度等级,用公式表示如下:

$$grade_{AC} = G_{AB} \times \frac{L_{AB}}{L_{AC}} + G_{BC} \times \frac{L_{BC}}{L_{AC}} \quad (1)$$

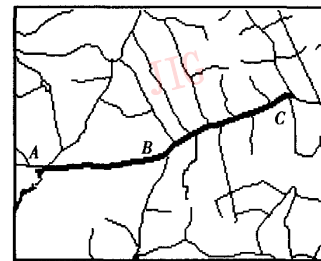


图 2 河渠的动态分段

Fig. 2 The dynamic segmentation of canal

式中, G 、 L 分别表示等级和长度, $grade_{AC}$ 表示河渠实体 AC 的整体密度等级。

数据库中的人工河网数据的存放并无彼此间的联系,且不像天然河系中河流之间存在明显的主支流关系。因此,本文认为不能像传统的道路网选取中采用的建拓扑、断链、构面的思路预处理数据,这样会使数据更加支离破碎。人工河网选取对河渠实

体进行整体的取舍,而不是在拓扑断链后对河渠实体中的某一条链进行取舍。故本文是在数据库中原有的人工河网数据上进行河渠空间关系的自动获取与识别,然后判断河渠与天然河流、湖泊的连接关系,进而自动获取和识别人工河网的密度,最后用面向对象的方法对获取以上信息的河渠实体进行封装。其过程如图 3 所示。

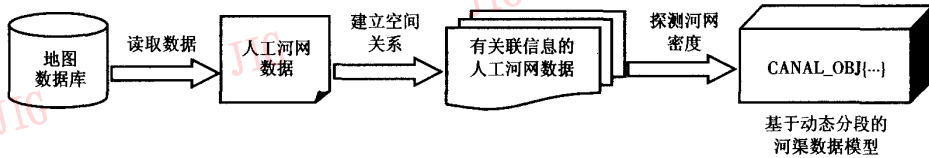


图 3 构建基于动态分段的河渠数据模型

Fig. 3 The data model of canal based on dynamic segmentation

3 人工河网空间知识的获取

3.1 人工河网空间关系的自动获取

由于人工河网中河渠间空间关系主要呈“┌”、“└”、“十”和“—*—”4 种形态,故这 4 种空间关系的获取只须探测出线要素之间的连接、相交关系即可。本文采用的具体步骤为:

(1) 建立线目标的栅格索引,依次判断每条线目标通过哪些栅格,并建立栅格与线的关联关系。

(2) 通过目标的栅格索引,只需判断每个栅格内的线目标是相交、相连(即“—*—”形)还是相离。若相交,首先判断是否为“十”相交,若不是则进而用参考文献[7]中的原理判断是“┌”形相交还是“└”形相交,并将相交的线的 ID 互相记录。

在判断两条线目标是否互相交之前,可以利用线目标之间的 MBR(外接矩形)是否重叠来排除不相交的情况,从而减少计算量。

(3) 获取人工河渠和天然水体(如河流、湖泊)之间的关系。本文采取的基本思路是用每条河渠的首、终结点分别进行搜索,看该点是否落在天然水体的范围内。为减少计算量,若点落入该要素的 MBR 内,才进一步判断点是否落在线目标或者面目标的外围轮廓线上。

3.2 人工河网密度的自动探测与识别

从图 1 中可以看出,不同区域的人工河网密度是有差异的,选取前必须首先识别出这种差异,并使之成为河渠实体的一种属性。传统的研究网络密度

的方法是基于图论原理构建网络的空间拓扑关系,计算网眼面积大小,根据网眼面积来判断网络的疏密程度。本文则提出了一种基于栅格的网络密度探测方法,其具体步骤是:

(1) 建立线目标的栅格索引,并将线目标分割存储,记录每个栅格内线目标的总长度 Len_i ;

(2) 计算该栅格的线目标密度 (km^2/km^2): $d_{ij} = Len_{ij}/S$ (Len_{ij} 为第 i 行,第 j 列的栅格的线目标总长度, S 为网格面积);

(3) 根据线目标密度值的大小将栅格进行分级,即令栅格中 $d_{max} = 1$,则第 i 行、第 j 列的栅格的相对密度值为 $D_{ij} = d_{ij}/d_{max}$ 。

(4) 根据图面情况将 D_{ij} 在 0~1 之间的取值分为若干级,得到每个栅格的线目标密度等级 G_{ij} ,将该等级数据按照线目标动态分段的思想存入经过该栅格的河渠实体的数据结构 CANAL_OBJ 中。

本文认为将图 1 的整个区域分为 4 个密度等级比较合适,表 1 是实验数据密度分级的指标,其分级结果如图 4 所示,能够较好地区分实验区域的密度差异。

表 1 密度分级指标

Tab. 1 The density classification of value D_{ij}

等级	D_{ij} 的取值范围
1	[0.0, 0.05)
2	[0.05, 0.31)
3	[0.31, 0.54]
4	[0.54, 1.0]

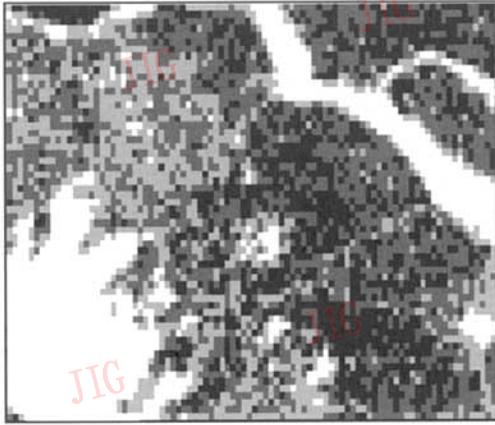


图 4 图 1 数据按线目标密度分级(4 级)
Fig. 4 The density classification of Fig. 1

4 基于“知识决策”的人工河网选取

4.1 必选河渠的确定

根据编绘规范,本文在河渠选取之前,首先依据编绘规范和制图区域的实际情况,确定一批必选河渠,然后整个人工河网的选取以这些必选河渠为骨架,在顾及河网密度差异和整体网络特征的情况下,利用基于知识的决策模型进行选取。必选河渠的确定主要依据以下原则:①图上长度大于 4mm 的“主要”河渠;②图上长度大于 4mm,且两端都与天然水体相连的河渠;③图上长度大于 20mm,且关联河渠数量不少于 5 的次要河渠。

4.2 人工河网选取的多准则决策模型

因河渠的选取须顾及多方面因素,在智能化方法中^[4-6],比较适合用多准则决策模型中的最大隶属度带权平均规划法对河渠进行较全面的评价后进行取舍^[7]。

(1) 确定决策目标集 $O = \{ \text{长度、主次属性、流域河网密度、与天然水体的连接情况} \}$

决策目标集中,长度用河渠的长度值 len 表示;主次属性(imp)的量化可分为“主要”和“次要”分别用 2 和 1 表示;河渠整体所经流域的河网密度可由式(1)扩展为

$$grade = \sum_{i=1}^n (G_i \times (len_i / len))$$

计算得到(G_i, len_i 分别为河渠实体不同分段的密度等级和长度, len 为河渠的总长度);与天然水体的连

接情况,根据连接不同的天然水体所表现出来的意义不同,给予大小不同的量化值,如表 2 所示。

表 2 河渠与天然水体连接情况量化表

Tab. 2 The sign of canal with river and lake

与天然水体的连接情况	量化值
两端都没有连接	1
只有一端连接天然水体	2
两端都连接河流或者湖泊	3
两端分别连接河流和湖泊	4

结合图 5 介绍该模型的建立过程。假设从数据中判断出河渠 AB 属于必选河渠,而与其相关联的 8 条河渠属于备选河渠,也就是决策的方案集 $X = \{ \text{必选河渠的关联河渠集合} \}$ 。结合决策的方案集 X 中每个备选方案的决策目标属性,可以得到河渠选取的决策矩阵

$$Y = \begin{bmatrix} len_1 & len_2 & \dots & len_n \\ imp_1 & imp_2 & \dots & imp_n \\ grade_1 & grade_2 & \dots & grade_n \\ link_1 & link_2 & \dots & link_n \end{bmatrix}$$

式中 $len_i, imp_i, grade_i, link_i$ 分别表示决策方案集中第 i 条河渠的长度、主次属性、流域平均密度等级以及与天然水体的连接属性等。然后用极小值标准化的方法转化为相对隶属度矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ r_{31} & r_{32} & \dots & r_{3n} \\ r_{41} & r_{42} & \dots & r_{4n} \end{bmatrix}$$

其中, $r_{1i} = \frac{len_i}{\max_{j=1}^n len_j}, r_{2i} = \frac{imp_i}{\max_{j=1}^n imp_j}, r_{3i} = \frac{grade_i}{\max_{j=1}^n grade_j}$,

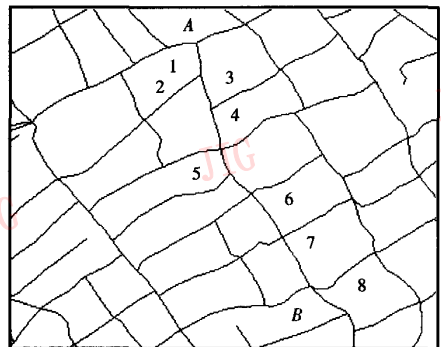


图 5 河渠选取多准则决策模型的建立

Fig. 5 The multiple criteria decision model of canal selection

$$r_{4i} = \frac{link_i}{\max_{j=1}^n link_j}, i = 1, \dots, n。$$

再结合目标集的权重组成权向量 $W = (w_1, w_2, w_3, w_4)$, 便可得到最终的决策向量 $D = W \times R$ 。将 $D = \{d_1, d_2, \dots, d_n\}$ 中的 n 个元素按从大到小的顺序排列, 便得到与必选河渠 AB 相关联的备选河渠的综合评价。

(2) 人工河网的取舍

上述模型中考虑了长度、主次属性、流域河网密度、与天然水体的连接情况这 4 个因素, 但模型中尚未讨论如何保持河渠的网络特征, 因网络特征很难量化到模型中, 为此本文提出如下解决方法。

① 求得决策向量 D 中最小的一个元素 d_{min} , 并找到它所对应的备选河渠 i_{min} 。显然, 第 i_{min} 条河渠是必须要舍去的。

② 判断 i_{min} 的奇偶性, 将备选河渠中序号与 i_{min} 奇偶性相同的河渠取出来, 并根据它们在决策向量 D 中所对应的元素 d 由小到大排序, 设为 $D' = \{d_{min}, d_{j1}, d_{j2}, \dots, d_{jm}\}$ 。

③ 根据经典的开方根规律公式 $n' = n \sqrt{(M/M')^x}$ (n 为原图上地物数, n' 为选取后地物数, M 为原图比例尺分母, M' 为目标比例尺分母, x 为选取等级), 可以计算出需要删除河渠的数量 $n - n'$ 。

④ 从 $d_{min}, d_{j1}, d_{j2}, \dots, d_{jm}$ 中选出前 n' 个, 它们所对应的备选河渠就是应当舍去的河渠。

例如图 5 中有 8 条备选河渠, 假设河渠 3 所对应的决策元素 d 最小, 那么对奇数序号的河渠根据 d 的大小进行排序, 设为 $\{3, 5, 7, 1\}$ 。根据计算, 8 条备选河渠中需要删除 2 条, 因此可将河渠 3、5 删除, 其余的保留。这样不仅顾及了选取数量, 而且利用河渠序号的奇偶做到了间隔选取, 从而较好地保持了人工河网的网络特征。

5 实验结果及结论

根据本模型, 从理论上可知影响最终选取结果的因素有: ① 权向量 W 和决定决策向量 D 中各元素 d 的排序。② 开方根规律公式中的选取等级 x , 影响选取的数量。③ 必选河渠的数量和分布, 影响选取的数量和分布。④ 是否使用 i_{min} 的奇偶性作为间隔选取因子, 影响人工河网网络特征的保持。

本实验中, 据统计河渠总数为 5 934 条, 依据规范必选河渠的数量为 1 123 条。实验发现, 使用间隔

选取因子时, 因为进行了控制隔一条选一条, 故 x 的大小对选取数量几乎没有影响, 这是很容易理解的。而在未使用间隔选取因子的情况下, 依照开方根公式计算的指标, 对排序后的决策向量 D' 中的元素从小到大进行选取, 此时 x 的影响就凸显出来。图 6 使用了间隔选取因子, 选取结果为 3 609 条, 选取率为 60.8%; 图 7 是没有采用间隔选取因子且 $x = 3.5$ 的选取结果, 选取数量为 3 350 条, 选取率为 56.5%。



图 6 使用间隔选取因子的实验结果

Fig. 6 The result of the distance between canals be considered



图 7 未使用间隔选取因子的实验结果

Fig. 7 The result of the distance between canals not be considered

实验中还发现选取数量的多少对保持网络结构特征有很大贡献, 这即是图 8 中 (数量为 3 859, 选取率为 65.0%) 在没有使用间隔因子 $x = 2$ 情况下, 整体网络特征的保持较图 6 和图 7 要好的原因。但选取数量过多使新编图的负载量过大, 也会影响读图的视觉效果。图 9 是仅按长度选取的实验结果, 显

然,在下方的湖泊群区域的处理和左上方网络结构的保持没有前 3 幅图的效果好,而且仅按长度选取也没有顾及到河渠本身的主次特征。表 3 是各种参数情况下的实验结果统计。

表 3 模型参数及结果

Tab.3 The parameter used in our model and the results

共同参数	α	间隔选取因子	结果		
			选取数量	选取率 (%)	结果图
河渠总数 5934	2.0	使用	3609	60.8	图 6
必选河渠 1123	3.5	未使用	3350	56.5	图 7
权向量 W : {0.6,0.2,0.1,0.1}	2.0	未使用	3859	65.0	图 8
按长度指标选取			2335	39.3	图 9



图 8 未使用间隔因子且 $\alpha=2$ 的实验结果

Fig.8 The result of $\alpha=2$ and the distance between canals not be considered

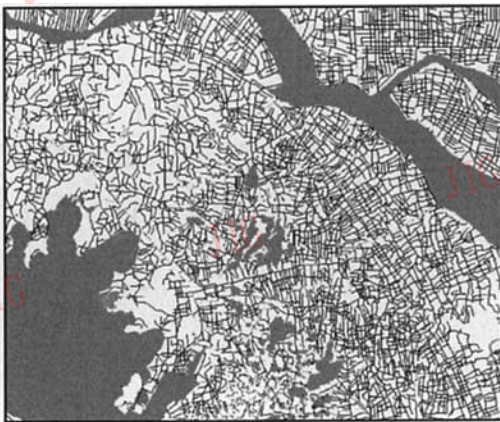


图 9 仅按长度选取的实验结果

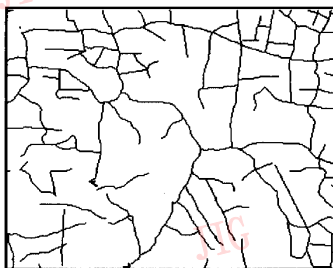
Fig.9 The result of selection only according to the length of canal

此外,权向量 W 对选取结果也有影响,但比较轻微,因为它主要影响到模型对决策向量中各个因素的排序,因此,其对总体数量没有多大影响,只影响细微处选取个体的差别。图 10 中的(a)和(b)分别是权向量直接赋值为 $W = \{0.6, 0.2, 0.1, 0.1\}$ 和权向量全部待定的情况下局部区域的选取结果。

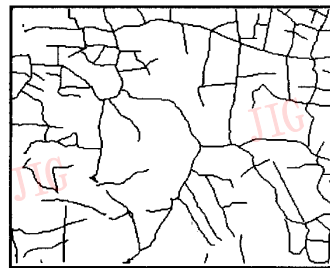
分析采用本方法选取后的结果图还可以发现,有些区域的图形结构特征的保持仍然不是很好,通过对数据的分析和原因的查找,其原因在于这些沟渠数据的采集极不规律,随意性很大,数据的组织很混乱。对这些区域内沟渠的取舍,还必须辅以人机交互的手段。

进行密集复杂人工沟渠网选取时,还要处理好沟渠和其他水系要素选取之间的关系。就本文的实验区域而言,存在大量湖泊群与沟渠交错分布的情况。湖泊群之间的许多沟渠都和长度选取指标相差较多,但因其与湖泊相连,因此在算法模型中有较大的选取可能性,有的甚至是必选的。因此本文认为在小湖泊分布密集的人工河网区,应先进行湖泊选取,然后进行人工河网的选取。

综上所述,本文建立的人工河网选取模型考虑了复杂人工河网综合中河渠的属性(主要、次要)、



(a) 权向量直接赋值



(b) 权向量全部待定

图 10 权向量取值方式不同时局部选取结果比较

Fig.10 The comparison in local results with different weight vector

长度、连通情况、河渠间的空间关系以及河网密度等五个方面的因素,以依据条件选取的必选河渠为基础,对每条备选河渠平等地进行综合评价,按照评价结果的选取可能性从小到大排序,然后根据间隔选取因子对河渠进行间隔选取,较好地保持了人工河网的网络特征。

参考文献 (References)

- 1 Wang Jia-yao, Fan Yi-ai, Han Tong-chun, *et al.* The Principle of Cartographic Generalization for General Map [M]. Beijing: Publishing House of Surveying and Mapping, 1992. [王家耀, 范亦爱, 韩同春等. 普通地图制图综合原理[M]. 北京:测绘出版社, 1992.]
- 2 Wang Jia-yao, Wu Fang. The Principle and Methods of Automated Cartographic Generalization for Digital Map [M]. Beijing: Publishing House of PLA, 1998. [王家耀, 武芳. 数字地图自动制图综合原理与方法[M]. 北京:解放军出版社, 1998.]
- 3 Zhang Qing-nian. Line generalization based on dynamic segmentation [J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2004, 43(2): 104 ~ 107. [张青年. 线状要素的动态分段与制图综合[J]. 中山大学学报(自然科学版), 2004, 43(2): 104 ~ 107.]
- 4 Guo Qing-sheng, Ren Xiao-yan. Intelligentize Processing of Geographic Information [M]. Wuhan: Publishing House of Wuhan University, 2003. [郭庆胜, 任晓燕. 智能化地理信息处理[M]. 武汉:武汉大学出版社, 2003.]
- 5 Gao Ji. Knowledge-based Intelligent Technology of Software [M]. Hangzhou: Publishing House of Zhejiang University, 2000. [高济. 基于知识的软件智能化技术[M]. 杭州:浙江大学出版社, 2000.]
- 6 Shao Li-xia, He Zong-yi, Ai Zi-xing, *et al.* Automatic generalization of river network based on BP neural network technique [J]. Geomatics and Information Science of Wuhan University, 2004, 29(6): 555 ~ 557. [邵黎霞, 何宗宜, 艾自兴等. 基于BP神经网络的河系自动综合研究[J]. 武汉大学学报(信息科学版), 2004, 29(6): 555 ~ 557.]
- 7 Tan Xiao, Wu Fang, Huang Qi, *et al.* A fuzzy decision approach for identifying master river and its application in river system construction [J]. Acta Geodaetica et Cartographica Sinica, 2005, 34(2): 154 ~ 160. [谭笑, 武芳, 黄琦等. 主流识别的多决策方法及其在河系结构化中的应用[J]. 测绘学报, 2005, 34(2): 154 ~ 160.]