

文章编号: 1001-1595(2007)01-0084-07

中图分类号: P208

文献标识码: A

多尺度道路网的距离匹配算法研究

陈玉敏¹, 龚健雅², 史文中³

(1. 武汉大学资源与环境科学学院, 湖北 武汉 430079; 2. 武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉 430079; 3. 香港理工大学土地测量与地理资讯学系, 香港 九龙)

A Distance-based Matching Algorithm for Multi-scale Road Networks

CHEN Yu-min¹, GONG Jian-ya², SHI Wen-zhong³

(1. School of Resource and Environment Science, Wuhan University, Wuhan 430079, China; 2. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; 3. Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China)

Abstract: Existing algorithms for geometry matching of multiple entities are short of meeting the requirements of multi-scale road network analysis due to algorithms' high complexity and long processing time. Based on the analysis of matching characteristics of poly-lines on road networks, this paper suggests the geometric similarities between poly-lines can be evaluated using the distances between poly-lines and nodes. As an alternative to a direct match of geometric similarities between poly-lines, this line node distance-based algorithm has a much lower computational complexity and the potential for efficiency improvement with a grid-based spatial index. Through a case study, we compared the distance-based algorithm with the traditional matching algorithm in computational efficiency and matching quality. We found the proposed algorithm is superior in both domains and will be able to meet the requirements for multi-scale road network analysis. We also conducted an analysis on the relationships between matching tolerance and precision using the least square method.

Key words: distance-based matching; geometry matching; multi-scale; road network

摘 要: 根据道路网折线的匹配特点, 提出基于格网索引的折线——结点距离匹配算法, 将复杂的折线与折线之间的几何相似度计算转换为求结点到折线距离的匹配方法, 降低了计算复杂度, 并通过建立格网索引来提高计算效率。在应用实例中, 采用曲线拟合的最小二乘法确定算法的匹配容差和匹配成功率之间的关系, 并通过与现有统计匹配算法比较, 可知该算法效率高且匹配成功率较理想, 能够满足多尺度道路网数据匹配的应用需求。

关键词: 距离匹配; 几何匹配; 多尺度; 道路网

1 引 言

道路网数据的多尺度特征从不同详细程度对道路实体进行表达或反映, 从而为用户提供从整体到局部、从抽象到具体的道路实体的管理和分析功能。目前, 世界上很多国家都建立了多尺度的道路网数据库, 例如我国建有 1: 100 万, 1: 25 万, 1: 10 万, 1: 5 万等不同比例尺的道路网数据库, 然而, 随着比例尺的变化, 道路实体发生着复

杂的数据和图形转换, 如聚合、简化、光滑等, 这使得同一条道路在不同比例尺的道路网中具有不同的几何信息和属性信息。如何确定道路要素在多尺度道路网数据库中的匹配关系和内在联系是有效处理多尺度道路网问题的关键。

对于数据集的匹配问题, 许多学者在地理信息系统、图像处理、交通等不同领域中进行了一定研究^[1]。研究重点主要是解决不同数据集的不一致性问题, 其不一致性可能是同一要素在不同

收稿日期: 2005-10-21; 修回日期: 2006-10-13

基金项目: 国家自然科学基金项目(40601082); 香港理工大学基金项目(G-YE13)

作者简介: 陈玉敏(1977-), 女, 山东栖霞人, 博士, 讲师, 主要从事网络地理信息系统和图像处理技术的研究工作。E-mail: ycym-lucky@hotmail.com

图上位置、几何形状、拓扑关系、属性数据等方面的不一致。最早的研究始于 20 世纪 80 年代中期,美国人口调查局有效集成了地质测量局的几何信息数据库和人口调查局的 TIGER 数据库,开发了世界上第一个地图自动合并系统,其匹配方法采用迭代算法,通过建立数据集中结点间的连接来确定不同地图之间的匹配,该算法只能处理 1:1 匹配关系^[6]。20 世纪 90 年代末期,各种改进算法相继出现,其中有学者提出一种基于统计特征的匹配算法,该算法采用关系匹配的思想,不仅仅考虑对象本身的匹配特征,同时考虑对象邻域的匹配情况,以提高匹配效率^[7]。到 21 世纪,研究工作由早期人机交互的半自动数据集匹配逐步转向全自动数据集匹配,有学者研究了三阶段式网络自动匹配算法,采用结点匹配、边匹配、线段匹配 3 种方式从结点、边和线段等不同方面进行匹配计算^[9]。

由研究现状可知目前匹配算法主要分为 3 类^[12]: ① 几何匹配,是通过计算几何相似度来进行同名实体的匹配; ② 拓扑匹配,是通过计算候选同名实体的拓扑关系度量作为匹配依据; ③ 语义匹配,是通过比较候选同名实体的语义信息。其中语义匹配是最有效的,如果两层数据中的两条道路具有相同的道路编码,则很容易判断是一对同名实体。但语义匹配在很大程度上依赖于数据模型以及属性数据类型等信息,而实际中来自不同数据库的多尺度数据往往缺少惟一标识的属性信息,无法判断,因此不得不采用拓扑匹配或几何匹配方式。而在多尺度道路网数据中,不同比例尺数据经过简化、退化等一系列的复杂图形变换,使得多尺度数据间的拓扑关系很难一致。如在 1:100 万道路网中一个交叉路口为四分支结构,但在 1:400 万道路网中该交叉路口合并为三支结构,以至拓扑匹配在多尺度道路网数据中无法应用。此时只能采用几何匹配进行计算。本文的研究重点就是基于距离匹配方式的多尺度道路网几何匹配计算方法。

道路网的几何匹配属于线状地物匹配范畴,存在多种情况,如图 1 所示^[7],分别表示为一条线对一条线的匹配(1:1),一条线对多条线的匹配(1:n),多条线对一条线的匹配(n:1),多条线对多条线的匹配(n:m),无匹配线(1:0),等等。

当前的几何匹配算法大多致力于自动匹配算法的研究与完善。对具有相似几何信息的 1:1 匹

配算法已经比较成熟,而对于几何信息、拓扑信息相差较大的 $n:m$ 匹配算法,由于缺乏可直接比较的标准,难以获得满意的匹配效率和精度,至今众多学者仍在寻找准确而高效的算法。在多尺度道路网中,随着比例尺的变化,要素的几何形状和表达内容都有较大差异,数据匹配属于较为复杂的 $n:m$ 匹配方式。本文针对多尺度道路网数据的特点,提出基于格网索引的折线-结点距离匹配算法以满足实际应用中 $n:m$ 匹配的快速计算需要。

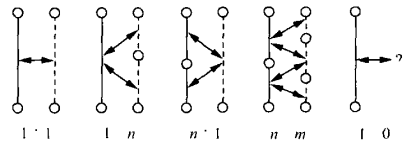


图 1 线状地物匹配情况^[7]

Fig. 1 The different situations of lines matching^[7]

2 多尺度道路网的距离匹配算法研究

几何匹配是根据地物几何形状的相似程度来进行匹配判断的,它是一个非常复杂的计算过程,主要问题在于地物的几何形状往往无法用一个精确值表示,而且来自不同数据源的数据其几何精度也各不相同,很难进行直接比较,因此在匹配计算过程中需要考虑的情况较多,匹配工作复杂,计算量大。本文所提出的距离匹配算法希望能有效解决这一难题。

2.1 距离匹配算法的基本思想

道路网属于线状地物,道路网中的每条道路通常采用道路拐点的连续坐标串记录,即采用折线表达,因而多尺度道路网的几何匹配算法实质上是计算折线状地物几何相似度的过程。折线与折线的几何相似度计算一般需要考虑到两折线之间的距离、两折线之间的夹角、两折线的长度等多种情况。另外,在多尺度道路网中很可能存在这种情况:1:400 万中的一条国道,在 1:100 万中分解为四条相邻的道路线段表示。此时需要计算 1:100 万中 4 条道路中任意一条道路与 1:400 万中该国道的某部分折线的相似度,即判断一条折线与另一条折线的某部分折线段的相似度,这大大增加了计算的复杂性。

考虑到一条道路折线可表达为一系列点坐标串集合,而求解“折线与折线之间的几何相似度”

远比计算“点到折线的距离”复杂,因此距离匹配算法的基本思想是将大比例尺数据(如 1:100 万)的道路网看作一个离散的“点集合”,小比例尺(如 1:400 万)的道路网仍然作为“折线集合”,将求解“折线与折线之间的几何相似度”问题转化为求解“点到折线的距离”;然后,通过寻找合适的距离匹配参量来确定大比例尺数据中满足要求的“匹配点集合”;最后,在“匹配点集合”中搜索可构成折线的“匹配线集合”。该“匹配线集合”就是所求的匹配结果。其可能存在多条匹配折线,从而有效解决了小比例尺道路网中的一条道路与大比例尺中的多条道路匹配的问题。这样可将复杂问题简单化,易于实现。

在进行多尺度道路网的距离匹配计算时,为了算法描述的一致性,假设遵循图 1 中所示的 1:n 匹配原则,即一对多原则,选择小比例尺数据作为已知的原始数据,大比例尺数据作为匹配目标数据,以保证原始数据中一条折线与目标数据中一条或多条折线进行匹配。对于 n:m 的匹配情况,则是遍历小比例尺数据中的 n 条折线,分别获得目标数据中一条或多条匹配折线,去除所有匹配折线中的重复部分,得到最终的 m 条匹配目标数据。

为了提高计算效率,可对大比例尺的目标数据建立等间距的规则格网索引。索引的建立具有以下几点优势:①确定目标数据中的各个结点的大致位置,规则格网索引的建立可以方便获得目标数据中各道路结点所在的格网索引,确定大致方位;②快速去除目标数据中的无关点集合,根据原始数据中一条道路的包围矩形得到相应格网,从而快速获得目标数据中对应于这些格网的点集合,去除无关的点集合,提高效率。

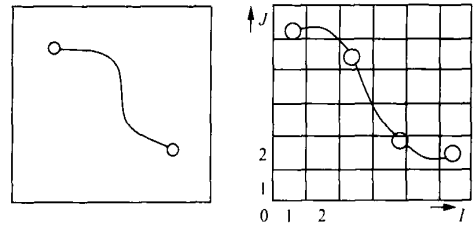
此外,该距离匹配算法需满足一个前提条件——多尺度道路网数据的坐标信息一致性。由于多尺度道路网数据来自不同比例尺的数据源,所采用的投影信息、坐标信息可能不一致,如小比例尺的数据可能是地理坐标或与之相当的等距离圆柱投影,而大比例尺的数据可为直角坐标系,投影一般为高斯投影。这将使得多尺度数据无法叠加进行匹配运算,因此在进行数据匹配算法之前必须进行坐标变换,以保证多尺度道路网的数据坐标信息一致性。

2.2 距离匹配算法的具体步骤

在多尺度道路网数据坐标信息一致的前提

下,基于格网索引的折线-结点距离匹配算法的具体步骤如下:

1. 确定距离匹配算法的原始数据和目标数据,按照 1:n 的匹配原则,将信息粗略的小比例尺数据作为原始数据,信息详细的大比例尺数据作为目标数据,如图 2 所示。



(a) 原始数据 (b) 用于匹配的目标数据
(a) Original data (b) Object data for matching

图 2 基于格网索引的折线-结点距离匹配

Fig. 2 The line node distance-based matching algorithm with a grid-based spatial index

2. 构建目标数据的等间距规则格网索引,如图 2(b) 所示。假设所有目标数据的范围为 $[(\min X, \min Y), (\max X, \max Y)]$, 则格网索引中 I, J 方向的每一格网大小 dx, dy 如式(1)所示,其中 m, n 分别为指定的 I, J 方向的格网数:

$$\left. \begin{aligned} dx &= (\max X - \min X) / m \\ dy &= (\max Y - \min Y) / n \end{aligned} \right\} \quad (1)$$

则目标数据中折线的任意一个结点(x, y),对应的格网坐标(i, j)如式(2)所示,其中[]表示取整符号:

$$\left. \begin{aligned} i &= [(x - \min X) / dx] \\ j &= [(y - \min Y) / dy] \end{aligned} \right\} \quad (2)$$

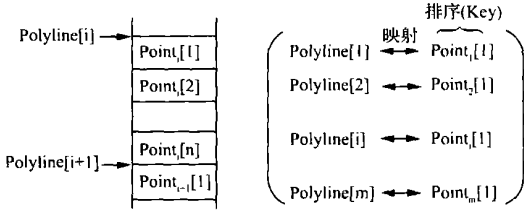
根据等间距格网索引规则,一对坐标(i, j)对应一个格网编号 h,如式(3)所示。这样就建立了目标数据中任意一个结点(x, y)与格网编号 h 之间的对应关系。

$$h = j \cdot m + i \quad (3)$$

3. 构建目标数据中结点与所属折线之间的对应关系。由于算法以目标数据的结点到原始数据中道路折线的距离为判断依据,当距离小于匹配容差 α 时(容差 α 根据实际问题可进行调整),可判断该结点满足匹配条件。此时需获知该结点在目标数据中所属折线,从而进一步判断该折线是否为匹配线。

目标数据中折线的结点是从 1 到 n 顺序记录的,当一条折线记录完后,记录下一条折线,如

图 3(a) 所示, 记录了目标数据中折线的数据结构。为了提高计算效率, 建立“折线和折线中第一个结点的映射”数组。该映射以“折线的第一个结点”作为关键字(Key), 并且按照“第一结点”的物理地址进行排序, 如图 3(b) 所示。



(a) 目标数据中折线的数据结构
(a) The poly-lines structure of object data
(b) 折线与折线中第一个结点的映射数组
(b) The mapping array of poly-line and first node in the poly-line

图 3 构建结点与所属折线之间的对应关系

Fig. 3 The relationships between nodes and poly-line

这样对于目标数据中任意一个结点, 可以快速查找位于映射数组中哪两个排序的“第一结点”之间, 从而判断该结点所属折线。例如, 已知结点 Pnt , 通过比较可知 Pnt 物理地址位于数组中 $Point_i[1]$ 和 $Point_{i+1}[1]$ 之间, 则可判断该结点 Pnt 是折线 $Polyline[i]$ 上的结点, 其折线的一个结点为 $Point_i[1]$ 。

4. 遍历原始数据(图 2(a)) 中一条折线 L 上每一线段(两个结点构成一条线段), 可得到在目标数据(图 2(b)) 中该折线 L 经过的所有格网, 从而获得目标数据中落入这些格网的结点集合 P 。

5. 在结点集合 P 中, 寻找所有与原始折线 L 距离小于匹配容差 α 的结点集合 P' 。求解方法: 对于集合 P 中的任意一结点 $P_i(X_0, Y_0)$, 求取该点到折线 L 每一线段的距离 D , 如图 4 和公式(4)所示, 取最小值 $\min D$ 作为结点到折线的距离。如果 $\min D$ 小于容差 α , 则判断 $P_i(X_0, Y_0)$ 属于集合 P' 。

$$\left. \begin{aligned} D &= |kX_0 - Y_0 + b| / \sqrt{k^2 + b^2} \\ k &= (Y_1 - Y_2) / (X_1 - X_2) \\ b &= Y_1 - k \times X_1 \end{aligned} \right\} \quad (4)$$

6. 计算匹配折线集合。通过结点与折线的映射关系数组(图 3(b)) 获得结点集合 P' 所对应的各条折线 l 。判断集合 P' 中属于折线 l 的结点个数

与折线 l 的总结点个数之比 κ 如式(5)所示, 其中 $inLine(pline)$ 表示在集合 P' 中属于折线 l 的结点个数, $Sum(pline)$ 表示折线 l 的总结点个数:

$$\kappa = inLine(pline) / Sum(pline) \times 100\% \quad (5)$$

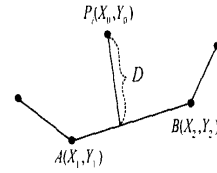


图 4 点到线段的距离

Fig. 4 The distance from point to line

当 κ 值大于或等于比例系数 β (β 可以根据实际问题进行调整, 如 $\beta = 80\%$) 时, 折线 l 与原始数据中折线 L 相匹配, 如式(6)所示, 其中 $pline$ 表示目标数据中的折线 l , $Matches(Line)$ 表示原始数据中折线 L 的匹配折线集合。计算所有与原始折线 L 相匹配的目标折线, 构成匹配折线集合。

$$\left. \begin{aligned} pline \in Matches(Line) & \quad (\kappa \geq \beta) \\ pline \notin Matches(Line) & \quad (\kappa < \beta) \end{aligned} \right\} \quad (6)$$

7. 遍历原始数据中所有待匹配的折线, 重复步骤 4~ 步骤 6, 获得各折线的匹配折线集合, 并去除匹配折线中的重复部分, 算法结束。

2.3 距离匹配算法的成功率

该距离匹配算法中有两个参量需根据实际问题进行调整: 匹配容差 α 和比例系数 β 。其中比例系数 β 反映折线 l 的结点落入集合 P' 的情况。最理想的是 $\beta = 100\%$, 折线 l 的所有结点在集合 P' 中, 则折线 l 一定为匹配折线。但由于不同比例尺数据间的复杂几何变化, $\beta = 100\%$ 的情况是非常少的, 一般有 80% 的点落入集合 P' 时即可判断为匹配折线。与比例系数 β 的取值相比, 匹配容差 α 取值尤为重要, 因为匹配容差 α 直接决定集合 P' 的大小和匹配成功率。匹配容差 α 反映了结点到折线 L 的距离范围, 对于不同的实际数据和匹配成功率要求, 其范围必定不同, 没有统一的表达式, 在下面应用实例中将结合实际数据详细分析。

值得注意的是在一些情况下并不需要对道路网所有地物进行匹配, 而只需对指定的一条道路进行匹配。对于这种单一对象的匹配问题, 容差 α 可以采用自动增量调节方式计算获得。首先容

差 α 给定一个任意小数, 进行匹配计算, 判断是否存在满足条件的匹配折线。如果没有时, α 迭代增加, 直到获得匹配折线。容差 α 的自动调节计算在单一对象匹配中非常重要, 能够计算得到 α 的确定值, 将很大程度上提高算法的准确性。

在比例系数 β 确定的情况下, 距离匹配算法的成功率 ε 与匹配容差 α 密切相关, 随着匹配容差 α 的变化而变化, 其函数关系在不同的实际应用中可能表现不同, 缺乏统一公式表达, 因此采用另一种可衡量值, 即匹配折线长度, 表示匹配成功率 ε , 如式(7)所示, 其中 $dic(c)$ 表示目标数据中正确匹配的折线长度, $dis(w)$ 表示目标数据中错误匹配的折线长度, $dis(sum)$ 表示目标数据中理想匹配的折线长度总和。

$$\varepsilon = (dis(c) - dis(w)) / dis(sum) \times 100\% \quad (7)$$

3 应用实例分析

对中国湖北省的一块区域 1: 100 万道路网和 1: 25 万道路网进行数据匹配计算。首先将两个不同比例尺的道路网数据进行投影、坐标统一变换后叠加显示, 如图 5 所示, 其中实线表示 1: 100 万道路网, 其数据量为 26 KB, 33 个对象; 虚线表示 1: 25 万道路网, 数据量为 524 KB, 344 个对象。

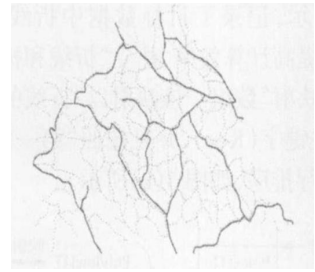


图 5 1: 100 万与 1: 25 万道路网叠加

Fig. 5 The overlapping map of 1: 1 000 000 and 1: 250 000 road network maps

采用距离匹配算法进行多尺度道路网的匹配计算, 将 1: 100 万道网作为原始数据, 1: 25 万道路网作为目标数据, 比例系数 β 设定为 80%。计算过程中, 匹配容差 α 的选择非常重要, α 过小无法找到匹配线段, α 过大可能会得到多余的匹配线段, 因此 α 的选择一定要适当。这里选取 11 组容差 α 值进行匹配结果比较, 容差 α 取值单位与地图实际坐标单位一致。容差 α , 匹配计算时间 t 和由公式(7) 计算匹配成功率 ε 的计算结果如表 1 所示。

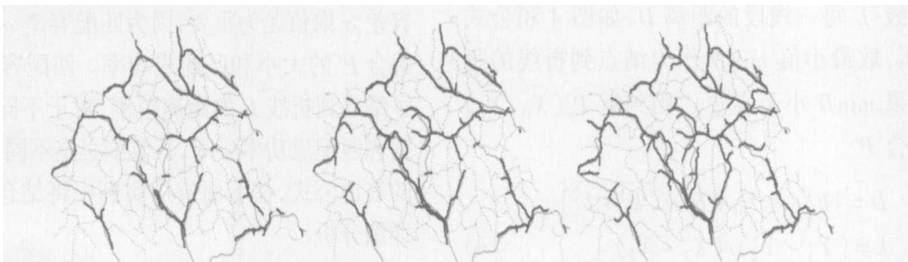
表 1 不同比例尺道路网数据匹配计算结果信息

Tab. 1 The matching result for different scale road networks map

| 容差 α | 1 000 | 1 100 | 1 200 | 1 300 | 1 400 | 1 500 | 1 600 | 1 700 | 1 800 | 1 900 | 2 000 |
|----------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 时间 t/ms | 34 250 | 34 250 | 34 260 | 34 270 | 34 274 | 34 282 | 34 292 | 34 300 | 34 304 | 34 320 | 34 328 |
| 成功率 $\varepsilon/\%$ | 87. 71 | 88. 84 | 90. 32 | 91. 06 | 91. 27 | 91. 12 | 89. 80 | 89. 41 | 87. 63 | 86. 54 | 85. 26 |

其中 3 组匹配结果如图 6 所示, 图中的粗黑线条代表从 1: 25 万道路网中选出的匹配 1: 100 万道路网的道路线。由图 6 可知, 当 $\alpha=1 000$ 时, 从 1: 25 万道路网中选出的匹配线段只能覆盖 1: 100 万道路网中的一部分, 还有一部分无法找到匹配线段, 匹配成功率为 87. 71%; 而当 $\alpha=$

2 000 时, 可以找到较多的匹配线段, 但同时也将一部分不匹配 1: 100 万道路网的小短线选进来, 需要做线段的剔除工作, 匹配成功率为 85. 26%; 当 $\alpha= 1 500$ 时, 搜索到较多的匹配线段, 且没有包含很多的冗余匹配线段, 匹配成功率为 91. 12%。



(a) 容差 $\alpha= 1 000$ (b) 容差 $\alpha= 1 500$ (c) 容差 $\alpha= 2 000$

图 6 算法中不同容差 α 对道路网数据的匹配影响

Fig. 6 The matching results for different matching tolerance α

匹配容差 α 应该如何选取, 才能使得匹配成功率 ε 最大? 将计算的 11 组数据绘制成散点图, 寻求匹配容差 α 和匹配成功率 ε 之间的关系, 如图 7 所示。将离散点连接成光滑曲线, 其形状类似于曲线函数。假定容差 α 和成功率 ε 之间满足二次曲线函数关系, 则 $\varepsilon = A\alpha^2 + B\alpha + C$, 根据 11 组样本值, 采用曲线拟合的最小二乘法, 计算参数 A, B, C , 如式(8)所示, 其中 $\varepsilon(\alpha_i)$ 表示通过二次曲线函数计算得到的函数值, 即 $\varepsilon(\alpha_i) = A\alpha_i^2 + B\alpha_i + C$, y_i 表示由表 2 得到的实际匹配成功率值, 最小二乘法的思想是使这两个值之差的平方和最小, 为计算方便可将各容差 α 除以 100, 使其与匹配成功率在同一数量级。求解得到参数 $A = -0.1804, B = 5.0998, C = 54.8896$, 可知匹配容差 α 和成功率 ε 的函数表达式为 $\varepsilon = -0.1804\alpha^2 + 5.0998\alpha + 54.8896$ 。求取曲线函数在 $[10, 20]$ 区间的最大值, 得: 当 $\alpha = 14.1347$ 时, $\varepsilon_{\max} = 90.9315$, 即当匹配容差 α 为 1 413. 47 时, 得到最大的匹配成功率 ε 为 90. 93%, 见图 7。这是通过二次曲线拟合计算出的最大匹配成功率, 而在该匹配容差 α 下的实际匹配成功率 ε 为 91. 32%, 误差仅为 0. 39%。对其他实际数据进行匹配时, 均可采用类似方法求得最优的匹配容差 α_0 。

$$\sum_{i=0}^{10} [\varepsilon(\alpha_i) - y_i]^2 = \min \quad (8)$$

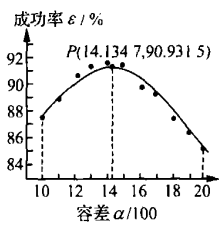


图 7 匹配容差 α 与匹配成功率 ε 的关系

Fig. 7 The relationships between matching tolerance α and precision ε

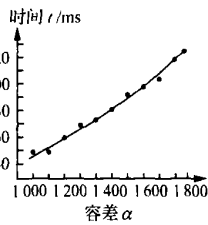


图 8 匹配容差 α 与时间 t 的关系

Fig. 8 The relationships between matching tolerance α and time t

算法的优劣除了与匹配成功率有关外, 还与计算效率密切相关。将表 2 中匹配容差 α 与时间 t 的关系绘制成散点图, 如图 8 所示。可见随着匹配容差 α 的增加, 计算时间 t 略有增加, 成单调递增关系。这是因为容差 α 增加, 计算的匹

配点集合 P' 增大, 获得的匹配线集合也在逐步增大, 所用计算时间相应增加。从实例中可知, 多尺度道路网的 33 个对象与 344 个对象的匹配, 在成功率最高的情况下, 即匹配容差 α 为 1 413. 47, 匹配成功率 ε 为 91. 32% 时, 计算时间约为 34. 28 s; 即使在成功率较低时, 即匹配容差 α 为 2 000, 成功率 ε 为 85. 26% 时, 计算时间约为 34. 33 s, 均在 1 min 以内。

为进行算法效率比较, 引用几何匹配中通常所采用的统计算法进行分析。Walter 和 Fritsch 采用的统计方法分别以道路的位置、长度、方向、夹角、邻接关系、道路缓冲区的范围等多个统计量为依据, 对可能的匹配线做出评价, 判断最合理的匹配线对^[7]。使用该统计算法对湖北省同一区域数据进行匹配计算, 33 个对象与 344 个对象的匹配时间约为 24 min 18 s, 匹配成功率为 92. 27%。这是由于计算每条道路的位置、方向、夹角, 尤其是道路的缓冲区比较费时, 在确定粗选匹配对后, 还需要根据各种统计量的组合判断置信度最高的匹配对, 其中循环判断比较的工作均消耗大量时间, 因此该统计算法所用时间是基于格网索引的折线- 结点距离匹配算法的 42. 5 倍。可见本文所提出的距离匹配算法, 在避免几何数据的各种复杂统计量计算情况下, 以结点到折线的距离为判断依据进行匹配计算, 效率是非常高的。此外, 距离匹配算法的最高匹配成功率 ε 为 91. 32%, 与统计算法得到的匹配成功率相差 0. 95%, 差异不到 1%。在综合算法效率和匹配成功率的情况下, 本文所提出的距离匹配算法整体优势是非常明显的, 可以满足多尺度道路网的 $n: m$ 几何匹配实际应用需求。

4 结束语

多尺度道路网的匹配问题在语义匹配和拓扑匹配无法进行计算时, 须采用几何匹配方式处理。而现有的几何匹配算法, 如基于统计特征的匹配算法, 在处理 $n: m$ 自动匹配时, 算法比较复杂, 计算时间长, 难以满足实际应用, 因此本文根据道路网中折线与折线的匹配特点, 提出了基于格网索引的折线- 结点距离匹配算法, 将复杂的“折线与折线之间的几何相似度”计算转换为简单的“结点到折线的距离”问题来进行匹配计算, 降低了计算复杂度, 同时采用建立等间距的格网索引方式来提高计算效率。从相同实际数据的应用分析可

见,该算法的匹配成功率为 91.32% 与统计算法的匹配成功率 92.27% 相差不到 1%,但计算效率提高了 42.5 倍。在避免复杂统计量的计算情况下,该算法在 33 个对象与 344 个对象的匹配中,用时不到 33 s,计算速度非常快。

该算法中匹配容差 α 的确定至关重要,直接关系到算法的匹配成功率和计算时间。采用曲线拟合的最小二乘法,对多组试验数据进行拟合可有效获得匹配容差 α 和匹配成功率 ε 之间的函数关系,从而确定最优匹配容差 α 和最高匹配成功率 ε 。随着匹配容差 α 的增加,计算时间 t 略有增加,当然,计算时间还与数据量相关,匹配数据对象个数的增加必将导致计算时间的增加。

综上所述,本文提出的基于格网索引的折线-结点距离匹配算法计算简单,效率高而且算法匹配成功率比较理想,可以满足道路网几何匹配的应用要求,有效解决了多尺度道路网的 $n:m$ 自动匹配问题。

参考文献:

- [1] DEVOGELE T, PARENT C, SPACCAPIETRA S. On Spatial Database Integration [J]. International Journal of Geographical Information Science, 1998, 12 (4): 335-352.
- [2] DUEKER K J, BUTLER J A. A Geographic Information System Framework for Transportation Data Sharing [J]. Transportation Research, Part C, 2000, 8 (1-6): 13-36.
- [3] KONCZ N A, ADAMS T M. A Data Model for Multi-dimensional Transportation Applications [J]. International Journal of Geographical Information Science, 2000, 16 (6): 551-569.
- [4] LAURINI R. Spatial Multi-database Topological Continuity and Indexing: A Step towards Seamless GIS Data Interoperability [J]. International Journal of Geographical Information Science, 1998, 12 (4): 373-402.
- [5] ROBERT F, FRANCOIS C. Map-image Matching Using a Multi-layer Perception: the Case of the Road Network [J]. ISPRS Journal of Photogrammetry & Remote Sensing, 1998, 53 (1): 76-78.
- [6] SAALFELD A. Automated Map Compilation [J]. International Journal of Geographical Information Systems, 1988, 2 (3): 217-228.
- [7] WALTER V, FRITSCH D. Matching Spatial Data Sets: a Statistical Approach [J]. International Journal of Geographic Information Science, 1999, 13 (5): 445-473.
- [8] WANG Y H, CHEN J, JIANG J. A Method for Linking Road-network Intersections in a Multi-scale Navigable Database [A]. International Archives of Photogrammetry and Remote Sensing[C]. [s. l.]: [s. n.], 2002. 511-515.
- [9] XIONG D. A Three-stage Computational Approach to Network Matching [J]. Transportation Research, Part C, 2000, 8 (1-6): 71-89.
- [10] XIONG D, SPERLING J. Semi-automated Matching for Network Database Integration [J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2004, 59 (1-2): 35-46.
- [11] ZHANG C. Towards an Operational System for Automated Updating of Road Databases by Integration of Imagery and Geodata [J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2004, 58 (3-4): 166-186.
- [12] ZHANG Q P, LI D R, GONG J Y. Map Conflation [J]. Bulletin of Surveying and Mapping, 2001, (7): 6-8. (张桥平,李德仁,龚健雅.地图合并技术[J].测绘通报,2001,(7):6-8.)

(责任编辑:雷秀丽)