

基于同步数据项组和碎分拓扑弧段时间 标记的时态地理数据模型^{*}

杜道生 舒 红

(武汉测绘科技大学测绘遥感信息工程国家重点实验室, 武汉市珞喻路 39号, 430070)

摘 要 从地理数据模型的定义出发, 阐述时间作为地理实体的基本特征及 GIS 中地理时间的模拟方法, 指出传统地理数据模型在地理环境变化快且应用强调细节变化分析的场合, 表现出不变数据的冗余存储、局部变化信息提取困难及地理实体的时间特征表现不明显等缺陷。基于现实世界中存在诸多同步变化的细小单元这一规律, 提出一种基于属性同步数据项组和空间碎分拓扑弧段时间标记的时态地理数据模型, 它降低了存储冗余且较好地表现了时间特征。

关键词 同步数据项组; 碎分拓扑弧段; 时态地理数据模型

分类号 TP 311. 12; P208

当今在地籍管理、环境模拟等领域, 历史状态的保存和恢复, 变化的跟踪和预测等应用要求非常强烈 (Langran Gail, 1992)。计算机大容量的存储器、高速的通讯和处理能力及遥感、全站仪、GPS 等时变地理数据采集手段的提高为时态数据库的建立提供了外部条件。传统地理数据模型适宜在地理状态变化缓慢的场合进行某一时刻的全局状态 (全局快照) 分析及宏观整体变化信息提取, 应用到地理状态变化频繁, 强调局部细节变化分析的场合则表现出明显的不变数据重复存储、局部变化信息提取困难、地理实体的时间特征表现不明显等缺陷。这些缺陷制约着时态地理数据库的建立和应用, 寻求克服这些缺陷的方案是 GIS 中时态地理数据模型研究的难点。时态地理数据模型是一种语义更完整的地理数据模型, 它不仅强调地理实体的空间和属性特征描述, 而且较传统地理数据模型更多地强调地理实体时间特征的描述。

1 地理实体的时间特征及地理时间的模拟方法

地理数据模型是人们为了一定的应用目的, 根据自己对客观地理世界 (或地理环境) 的认识, 以数字数据的形式建立起对客观地理世界 (或地理环境) 的模拟系统。对客观地理世界的模拟要求我们尽可能接近其本来面目, 即地理数据

模型必须强调现实的地理涵义 (或称语义)。客观地理世界是纷繁复杂、变化万千的, 无法也没有必要对其进行完全真实的模拟, 最好的办法是将客观地理世界看成是按一定地理特征区别开来, 同时又相互关联的地理实体集合。之所以说我们无法也没有必要对客观地理世界进行完全真实的模拟, 原因在于数据模型的建立受限于人们对客观地理世界的认识深度和用户具体的应用要求。实际的做法是人类尽最大的努力去认识客观地理世界, 从用户的应用目的出发, 对客观地理世界及其相关的事物进行一定范围一定深度的模拟, 即地理数据模型必须强调用户的要求。如地质学家看重数据模型对地壳形变的反映, 而水利部门要求数据模型对江河湖海水势情况的逼真模拟。人地系统是一相互作用的巨系统, 强调地理实体的本质特征和兼顾用户要求体现了人类揭示自然本质、服务于自身的对立统一的认识过程。此外, 数据模型是指以数字化形式数据建立起的客观地理世界模拟系统, 这说明除语义和用户要求外还必须强调模型在计算机上的有效实现。

任何地理实体都处在一定的地理时空坐标系里, 地理属性、地理空间和地理时间是地理实体 3 个最基本的特征。地理属性特征偏重地理实体质量和度量信息的描述, 地理空间特征偏重地理实体在地球表面及附近空间分布的描述, 地理时间特征则偏重地理实体时间尺度和时态关系的描述。地理时间特征可通过地理状态 (包括地理属

收稿日期: 1996-11-18 杜道生, 男, 55岁, 教授, 现从事地理信息系统研究。

^{*} 国家自然科学基金、国家教委霍英东教育基金、国家教委博士点基金资助项目, 编号 49671063

性状态和地理空间状态)的绝对时间标记和出现的相对次序来存储,地理实体的时变信息通过对不同的地理状态分析来提取。依据状态变化,人们进行地理事件或地理过程的解释。事件描述了一个产生结果的活动,而过程则不涉及结果。如“河水上涨”是一地理过程,而“河水淹没了房屋”则属一地理事件。地理时间模拟方法如图 1 所示。

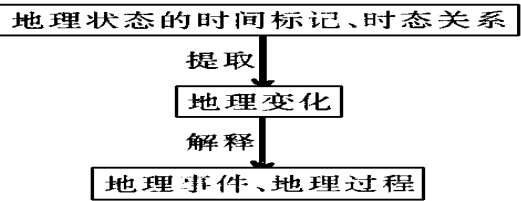


图 1 地理时间模拟方法

Fig. 1 Geographic Time Modeling Method

时间的本质为事件序列。地理时间以地理事件发生的频率来决定其时间尺度(或称时间分辨率)。最本质地刻画地理时间的方法是直接记录发生的地理事件序列(线性或分支序列),如图 2 所示为地理事件驱动的时态数据模型。由于真实事件的发现受限于人类改造自然和认识自然的深度,因此揭示引起地理状态变化的地理事件或过程有相当难度,单纯以地理事件或地理过程来对地理时间进行研究可归于地理数据分析或时空推理的范畴。地理实体时间特征描述主要有两种:一是以绝对时间记载地理实体某一状态的有效期,二是以先后等相对时间关系(又称时态关系)记载地理实体不同状态出现的次序,也就是说地理实体时间特征描述提供了地理实体有效状态的时间定位和不同状态间的时态关系分析两种功能。最常见的时态关系是时间上的前后关系,理论上把时间模拟成具有“ \leq (前后)”全序关系的实数 R 有理数 Q 或自然数 N 的集合 T 。设 $T = \{t_0, t_1, t_2, \dots, t_{\text{now}}, \dots\} \subseteq N$, 其中 t_{now} 对应当前时刻(时间点),两相邻时间点的距离。时间尺度主要取决于事物本身变化的频繁度和应用要求。实体状态的维持期可以时间点 t_i 和时间段 $t_i - t_j$ 表示,以集合表示时刻 $T_s = \{t_m\}$ 和时间段 $T_s = \{t_k | t_k \leq t_j, t_k \in T\}$, T_s 为 T 的子集 ($T_s \subseteq T$),其在状态 A 的维持期为 T_{s_a} ,在状态 B 的维持期为 T_{s_b} 。 T_{s_a} 和 T_{s_b} 定义了 2×13 种时态关系(舒红,陈军,1996)。

通常,人们对地理时间研究存在两种观念:一是强调地理实体的完整意义,解释地理实体存在生命期的时间构成关系及不同实体生命期的相对关系(从数学角度定义其为时态拓扑关系);二是

以地理事件为中心点,重点分析各种粒度的不同地理状态变化。我们认为研究时态地理数据模型既要强调地理实体的绝对时间位置和相对时间关系,还要强调地理状态的变化分析。地理事件序列直接表达了地理实体的时间语义,记录或分析时态关系不仅隐式反映了地理事件,而且更深层地表明了地理事件间的制约关系。地理实体变化主要有局部的变化(实体进化)和地理实体的存亡。地理实体的存亡由地理实体的本质特征引起,而本质特征常常为用户感兴趣或着重强调的特征,有可能是某种属性特征,也有可能是某种空间特征。

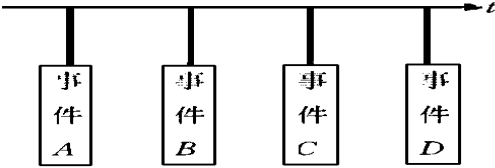


图 2 地理事件驱动的时态数据模型

Fig. 2 Geographic Event Driven Temporal Data Model

2 基于同步数据项组和碎分拓扑弧段时间标记的时态地理数据模型

理论上,时间线轴从过去向未来无限延伸,引起状态变化的事件不断发生,数据存储量将会无限增长。设计降低不变数据存储空间、自然地表现实体时间特征(尤其是时态关系)的数据结构及其有效操作是设计时态地理数据模型的基本指导思想。

时间标记是指在记录对象某一地理状态的同时,记录下该状态在现实世界中出现的有效时间(绝对时间位置)。时间标记的内容有标记两状态差值和标记新状态两种,这里采用标记新状态的办法。标记的时间元素有时间点、时间段和时态元素标记,这里采用时间段标记的办法。标记的单位大小有全局状态时间标记和局部状态时间标记。全局状态时间标记能记载全局整体变化效应,适宜于变化缓慢或存在大量同步变化特征项,应用强调某一时刻宏观数据联系及状态快速提取的场合,传统的 GIS 采用的即是全局状态时间标记。局部状态时间标记则更能体现局部细节变化,避免不变数据的重复存储,适宜于变化频繁或各特征项异步变化突出且应用强调变化分析和时态信息快速提取的场合。以传统 GIS 全局状态时间标记的数据模型去对付局部状态时间标记的场合,

则局部状态时间标记数据模型的优点将成为其缺点,即不变状态数据的冗余存储,不便于局部细节变化的分析,时间维信息表现不明显等。时态数据模型强调地理实体时间语义的保持,时间维信息的提取和局部变化分析,应采用局部状态时间标记方法。

时间标记的最小单位为属性数据项和空间坐标点。事物现象的同步变化是现实世界中的一种客观规律。因此,在研究范围内,采用将同步变化的若干最小单位归为一组并附上一个时间标记的办法。具体说来,属性状态里同步变化的若干数据项归为一组——同步数据项组,空间状态里一定时间范围里同步变化的若干坐标点归为一组——

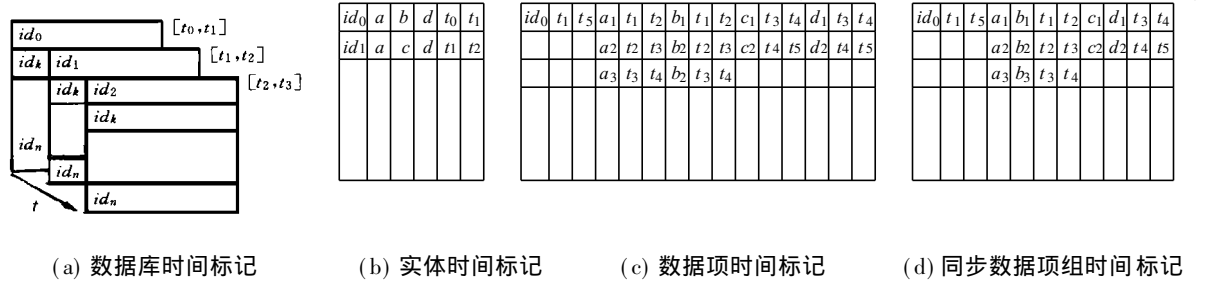


图 3 属性数据的时间标记对象

Fig. 3 Timestamped Units of Attribute Data

间标记(图 3(a))中,稍有某一实体某一数据项的值发生变化($id_0 \rightarrow id_1 \rightarrow id_2$)就重新存储一新数据库,显然存在不变实体数据(id_k, \dots, id_n)的冗余存储。实体时间标记(图 3(b))中,实体数据项值稍有变化(id_0 的某一数据项值由 b 改变为 c)就另存储一新的实体(id_1),两实体不变数据项值(a, d)的重复存储是一种存储冗余。数据项时间标记(图 3(c))中,每一数据项的各个值都单独带上时间标记,对于大量同步变化的数据项,相同时间标记的存储同样是一种冗余。由于变化发生的最小单位是数据项,实体内部属性数据项间存在一对一的同步变化关系,建议将时间标记到同步数据项组上,组内的数据项数可为一个或若干个。几种属性数据时间标记方法性能比较如下:同步数据项组时间标记(图 3(d))较其他方法更节省存储空间;从强调快速提取某一时刻全局信息角度看,优良次序为 (a) \rightarrow (b) \rightarrow (d) \rightarrow (c);从强调提取局部变化的详细程度、速度和时间维信息表现明显度看,优良次序为 (c) \rightarrow (d) \rightarrow (b) \rightarrow (a)。权衡几种方法,(d)较其他方法更适合建立时态地理数据模型。设描述实体的数据项有 User- id $X_2 X_3 X_4 X_5 X_6 X_7 X_8 X_9 X_{10}$,其中 User- id 数据项为实体标识, $X_2 X_3$ 为一同步数据项组, $X_8 X_9 X_{10}$ 为另一同步数据项组,则实体数据结构表示为:

User- id $T_{s_0} T_{e_0} \{X_4 T_{s_4} T_{e_4}\} \{X_5 T_{s_5}$

碎分拓扑弧段时间标记,可以达到既体现地理状态的局部变化又最大限度避免不变状态数据冗余存储的效果。它在实体内全部单元值变化同步时演化为实体时间标记,而在实体内各数据单元变化节奏各不相同同时演化为最小单元时间标记(属性项时间标记,坐标点时间标记)。

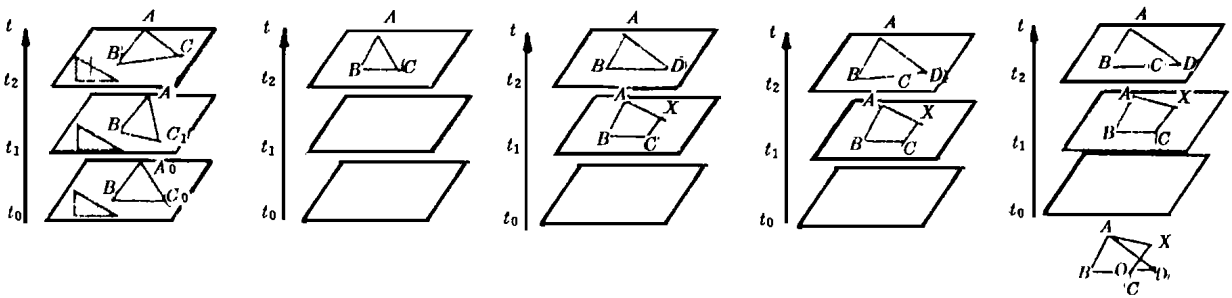
地理实体的地理状态由属性状态和空间状态组成。所谓的全局状态或局部状态也只是一个相对的概念。下面将分析比较各种属性状态时间标记和 point-arc-polygon 类型矢量结构空间状态时间标记方法。仅考虑属性数据,将时间标记的对象分为数据库、实体和数据项 3 个层次(图 3)。

数据库(或研究范围全体实体某一状态)时

$$T_{e_5} \{X_6 T_{s_6} T_{e_6}\} \{X_7 T_{s_7} T_{e_7}\} \{X_2 X_3 T_{s_2} T_{e_2}\} \{X_8 X_9 X_{10} T_{s_8} T_{e_8}\}$$

考虑矢量结构的时空状态时间标记,将时间标记的对象分成数据库、实体、弧段、坐标点 3 个层次,见图 4。

数据库时间标记(图 4(a))中,稍有某一实体的空间分布状况发生变化(多边形 $A_0BC_0 \rightarrow$ 多边形 $A_1BC_1 \rightarrow$ 多边形 ABC),就另存一新的数据库,显然存在不变实体(polygon)空间数据的冗余存储。坐标点时间标记(图 4(b))中,同一时刻弧段上所有坐标点具有相同时间标记,相同 t 值的存储是一种冗余。如 ArcAB: $(x_1, y_1, t_s, t_e) (x_2, y_2, t_s, t_e) \dots (x_n, y_n, t_s, t_e)$ 。地理实体时间标记(图 4(c))中,某一实体发生局部的形状改变(ArcAXC \rightarrow ArcAC),就重新另存一新的实体,实体不变弧段存储(ArcAB 与 ArcBC,在 t_s, t_e 时刻的双份存储)属冗余。弧段时间标记(图 4(c'))中,弧段稍有改变(ArcBC \rightarrow ArcBD, ArcAXC \rightarrow ArcAD),就另存一份新的弧段。其实,弧段上只有一小段发生改变(ArcBC 增加了直线段 CD 如图 4(c')),仍然存在弧段不变部分坐标点数据的冗余存储。采用将时间标记在同步变化的坐标点集合——碎分弧段上(图 4(d)),具体方法为将各个不同时刻的地理实体空间分布投影到某一时刻空间平面上,进行虚拟求交(没有实在物理意义的求交)即可得到碎分弧段,在碎分



(a) 数据库时间标记 (b) 坐标点时间标记 (c) 实体时间标记 (d) 弧段时间标记 (e) 碎分弧段时间标记

图 4 空间数据的时间标记对象

Fig. 4 Timestamped Units of Spatial Data

弧段上记录它存在的有效时间。这种方法对碎分弧段只进行一次存储,不变坐标点数据只进行了一次存储,同时由于碎分弧段上记录了有效时间,可通过时间标记检索出构成某时刻地理实体实际弧段的碎分弧段集合来。几种空间数据时间标记方法性能比较如下:碎分拓扑弧段时间标记更节省存储空间;从强调快速提取某一时刻的全局空间信息角度看,优良次序为 (a)→(c)→(d)→(b);从强调提取空间局部细节变化的详细程度、速度及时间维信息表现明显度看,优良次序为 (b)→(d)→(c)→(a)。综合比较,(d) 较其他方法更适合建立时态地理数据模型。基于 point-arc-polygon 矢量数据结构,提出如下体现空间变化的数据结构及其 4 个基本的时变操作,这里的 point- id line- id 和 poly- id 相当于属性数据结构里的 User- id (用户标识码),Arc#、Sd#、Ed#、Lp#、Rp# 分别为碎分弧段及其起始节点、左右多边形的标识码。Ts、Te 为 Lp# 的有效时间段 [Ts, Te] 标记。

基本时间标记单元——碎分拓扑弧段:

Arc Arc# Sd# Ed# (x0,y0)(x1,y1), ..., (xn,yn)
Ts Te {Lp# Ts1 Te1} {Rp# Ts2 Te2}

简单点、线、面地物

Point point- id (x,y) Ts Te
Line line- id {line# Arc# Arc2# , ..., Arcn# Ts Te}
Polygon poly- id {polygon# Arc# Arc2# , ..., Arcn# Ts Te}

4 个基本的时变操作:

独立点的增加,弧段增加,弧段分裂和弧段合并。

注意:时态地理数据模型语义上只有过时的概念,没有删除操作。

对于用户方进行的一些地理实体时变操作,如独立地物的增加、独立地物的删除、实体的形态改变、实体的大小改变、实体的方向改变、实体的

位置改变、实体的集合构成替代、实体间的拓扑关系改变等操作均可以通过 4 个基本的时变操作实现。下面,给出一个地理实体分裂的例子(图 5),同时说明相应的分裂操作和文件存储情况。

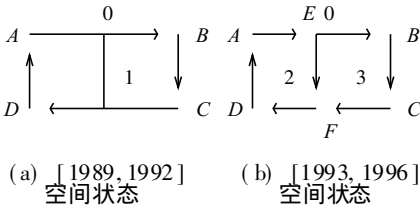


图 5 地理实体分裂

Fig. 5 Division of Geographic Entity

图 5 (a) 时刻计算机内弧段文件:

ArcAB# NodeA# NodeB# (xa0, ya0) (xa1, ya1)
... (xai, yai) 1989 1992 { Polygon0# 1989 1992 } { Polygon# 1989 1992 }
ArcBC# NodeB# NodeC# (xb0, yb0) (xb1, yb1)
... (xbj, ybj) 1989 1992 { Polygon0# 1989 1992 } { Polygon# 1989 1992 }
ArcCD# NodeC# NodeD# (xc0, yc0) (xc1, yc1)
... (xcn, ycn) 1989 1992 { Polygon0# 1989 1992 } { Polygon# 1989 1992 }
ArcDA# NodeD# NodeA# (xd0, yd0) (xd1, yd1)
... (xdn, ydn) 1989 1992 { Polygon0# 1989 1992 } { Polygon# 1989 1992 }

图 5 (a) 时刻计算机内地理实体文件:

{ Polygon1# ArcAB ArcBC ArcCD ArcDA 1989 1992 } Polygon1- id

地理实体的分裂 (Polygon1- id→ Polygon2- id, Polygon3- id) 由基本时变操作:弧段分裂 (ArcAB→ ArcAE+ ArcEB, ArcCD→ ArcCF+ ArcFD) 和弧段增加 (ArcEF) 来完成。内部实现操作为删除原弧段文件的 ArcAB 和 ArcCD,增加 ArcEF ArcAE ArcEB ArcCF ArcFD,且 ArcAE ArcEB ArcCF 和 ArcFD 的有效时间为 [1989, 1996],而非 [1993, 1996]。

图 5 (b) 时刻计算机内弧段文件:

ArcAE# NodeA# NodeE# (x_{a_0}, y_{a_0}) (x_{a_1}, y_{a_1})
... (x_{e_0}, y_{e_0}) 1989 1996 { Polygon0# 1989 1996} { Poly-
gon# 1989 1992} { Polygon2# 1993 1996}
ArcEB# NodeB# NodeB# (x_{e_0}, y_{e_0}) (x_{e_1}, y_{e_1})
... (x_{a_i}, y_{a_i}) 1989 1996 { Polygon0# 1989 1996} { Poly-
gon# 1989 1992} { Polygon3# 1993 1996}
ArcBC# NodeB# NodeC# (x_{a_i}, y_{a_i}) (x_{b_1}, y_{b_1})
... (x_{b_j}, y_{b_j}) 1989 1996 { Polygon0# 1989 1996} { Poly-
gon# 1989 1992} { Polygon3# 1993 1996}
ArcCF# NodeC# NodeF# (x_{b_j}, y_{b_j}) (x_{c_1}, y_{c_1})
... (x_{f_0}, y_{f_0}) 1989 1996 { Polygon0# 1989 1996} { Poly-
gon# 1989 1992} { Polygon3# 1993 1996}
ArcFD# NodeF# NodeD# (x_{f_0}, y_{f_0}) (x_{f_1}, y_{f_1})
... (x_{c_m}, y_{c_m}) 1989 1996 { Polygon0# 1989 1996} { Poly-
gon# 1989 1992} { Polygon2# 1993 1996}
ArcDA# NodeD# NodeA# (x_{c_m}, y_{c_m}) (x_{d_1}, y_{d_1})
... (x_{d_n}, y_{d_n}) 1989 1996 { Polygon0# 1989 1996} { Poly-
gon# 1989 1992} { Polygon2# 1993 1996}
ArcEF# NodeE# NodeF# (x_{e_0}, y_{e_0}) (x_{e_1}, y_{e_1})
... (x_{f_0}, y_{f_0}) 1993 1996 { Polygon3# 1993 1996} { Poly-
gon2# 1993 1996}

图 5 (b) 时刻计算机内地理实体文件:

{ Polygon1# ArcAE# ArcEB# ArcBC#
ArcCF# ArcFD# ArcDA# 1989 1992} Poly-
gon1.id
{ Polygon2# ArcAE# ArcEF# ArcFD#
ArcDA# 1993 1996} Polygon2.id
{ Polygon3# ArcEB# ArcBC# ArcCF#
ArcEF# 1993 1992} Polygon3.id

由上可以看出,碎分弧段的时间标记不仅可以
将地理实体各种历史状态恢复出来,方便地理
实体空间几何变化信息检索,而且它记录了起始
节点 Sd#、终止节点 Ed#、各个时期的左右邻接
多边形的内部标识号 Lp# 与 Rp#,存有多边形
各个时期的碎分弧段组成,相当于记录了实体空
间单元联接关系的时间标记,故称所标记的弧段
为碎分拓扑弧段,使常用面域定义、邻接和连通 3
种空间拓扑关系的变化信息的提取变得尤其容
易。图 6 为空间状态变化部分实例 (其中虚线和
实线同为碎分弧段),表明了地理实体的各种时
变操作可化解为若干基本时变操作。

物理实现方面有两种方案: 其一是将集合项
(如属性同步数据项组 $\{X^2 X^3 T_{s_2} T_{e_2}\} \{X^8 X^9 X^{10}$
 $T_{s_8} T_{e_8}\}$,时变空间状态数据集 $\{Lp\# T_{s_1} T_{e_1}\}$
 $\{Rp\# T_{s_2} T_{e_2}\} \{lin\# Arc\# Arc2\#, \dots, Arcn$
 $\# T_{s_3} T_{e_3}\} \{polygon\# Arc\# Arc2\#, \dots, Arcn$
 $\# T_{s_3} T_{e_3}\}$) 定义为新的数据类,以面向对象数据
库的机制加以实现; 其二是采用扩充关系数据库,

增加序列、元组和集合数据类型的非第一范式
NF²来实现。提时态数据库和静态数据库是从时
间语义角度讲,提面向对象数据库和扩充的关系
数据库是从计算机实现技术角度讲。面向对象方
法表现时态语义可采用历史版本机制实现,如何
利用低层次、部分时间标记对象 (对象或属性版
本) 构造高层次、复合时间标记对象是面向对象
数据库技术的一个难题。采用关系表格方法表现
时态语义必须突破关系数据库本身的规范化限
制。由于关系数据库是基于值、面向元组的,为了
得到一个实体的完整信息,常要在多个元组间做
若干代数运算 (连接、投影、选择、并、差等),必须
发展 NF²的时态关系代数和时态谓词演算。

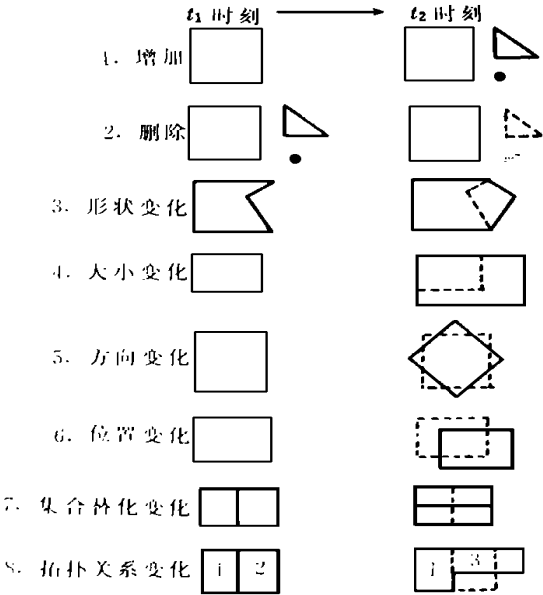


图 6 空间状态变化部分实例

Fig. 6 Some Spatial Status changes Examples

3 结束语

碎分拓扑弧段时间标记方法由于将地理实体的
空间数据进行多次人为的分裂,必然出现大量
碎片,从而带来地理实体历史空间数据重组的低
效率。为此,采取设立多个基态和由用户依据应用
要求给出分裂容忍值 (Split-tolerance)的措施来
控制弧段分裂过碎。对于相邻两弧段若其变化微
小到小于容忍值所确定的范围,则忽略变化,不进
行弧段分裂 (如图 7)。

本文仅从语义和结构上对时态地理数据模型
进行了一定的探讨。研究模型的数学定义 (时态
数据类型、时态地理实体、时空关系、对象代数和
逻辑描述)及物理实现 (时态数据存储、针对频

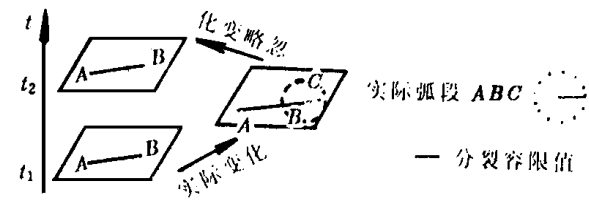


图 7 控制弧段碎分的分裂容限值

Fig. 7 Split-Tolerance Controlling Segemented Arc

常用数据和非常用数据分别采用本地机/快速存

储介质和远程服务器 光盘存放的组织存放策略、时空联合索引、时态查询操作和推理分析机制、时空数据可视化等)是实现时态 DBMS 必须解决的问题。

参 考 文 献

1 Langran G. Time in Geographical Information System. Taylor & Francis Ltd, 1992

2 舒红,陈军.地理数据模型中的时间语义.见:第二届全国地理信息系统年会论文集.北京,1996.

A Temporal Geo-data Model with Timestamp
on the Group of Synchronous Changing Data Items
and the Segmented Topological Arc

Du Daosheng Shu Hong

(National Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing,
W TUSM, 39 Luoyu Road, Wuhan, China, 430070)

Abstract According to definition of geo-data model, it has been analyzed that time is a basic property of geographic entity and how to study geographic time in GIS. In the case of reality geographical environment changing fastly and application focusing on analysis of local change, current geo-data model have many flaws such as redundant storage of unchanged data, difficult extraction of temporal information, implicit representation of temporal property of geographic entity. Based on the law of existing many synchronous changing small units, a temporal geo-data model with timestamp on the group of synchronous changing data items and the segmented topological arc is presented, which reduces redundant storage and shows us apparent temporal property of geographic entity.

Key words synchronous changing data items; segmented topological arc; temporal geo-data model

《武汉测绘科技大学学报》编辑委员会

名誉主任: 宁津生

主 任: 李德仁

委 员: 刘经南 王新洲 朱灼文 晁定波 张正禄

郑肇葆 龚健雅 舒 宁 许云涛 张祖勋

毋河海 刘耀林 李 霖 杜道生 朱元泓

詹庆明 李清泉 赵茂泰 梁荫中 柳建乔

主 编: 李德仁 (兼)

副 主 编: 柳建乔