DOI: 10.13203/j.whugis1996.02.016

第21卷第2期 武汉测绘科技大学学报 Vol. 21 No. 2
1996年6月 Journal of Wuhan Technical University of Surveying and Mapping June 1996

N1NF时态数据库及其更新操作

黄明智 张祖勋
(河海大学测量工程系,南京市西康路1号,210098)

摘 要 作者将"调查时间"引进时态数据库,使得关系运算中对于"当前时间"的处理更准确、更完善,进而定义了一种基于关系运算的更新操作机制;对现有的时态数据理论作了其它扩充、简化和改进,使结果更加简明实用。

关键词 时态数据库;非第一范式;数据更新
分类号 TP311.13

传统的关系模型受1NF(First Normal Form,第一范式)的制约,不宜于处理以时间为参照的非表格化数据。近10余年来人们开始了N1NF(Non-1NF,非第一范式)数据模型的研究,以支持非商业应用的大而复杂的结构目标[1~3]。其中有关时态数据库的研究占了较大比重,以文献[4,5]较有代表性:文献[4]引进了时态元素(temporal element)和时态赋值(temporal assignment)的概念,建立了一种有特色的时态模型;文献[5]对前者作了改进,在关系运算中对不确定时间值"now"(表示当前时间)进行了初步处理。

为了深化时态数据库的研究,同时将研究结果实用化、工程化,全面严格地讨论时态数据的组织、表达和操作,尤其是数据的更新(renewal)操作,是十分必要的,但文献[4,5]对此做得都还很不够。在本文中作者引进了调查时间作为用户定义(user-defined)时间,使得关系运算中对于"now"的处理比文献[4]更准确、更全面;进而作者为时态数据库建立了一种基于关系运算的更新操作机制。

作者还在文献[4,5]的基础上加强或增加了时态区间和广义属性集的概念,因而既省去了时态元素的概念,又弱化了时态赋值的理论限制;并且应用了目标的生命区间,强化了时态操作的一致性。

1 时间标记和时态区间

1.1 时间标记和生命区间

我们同时采用3种时间标记:事实时间(或称有效时间,valid time[4])——目标在现实世界中的产生(birth)时间 t_b 和消亡(death)时间 t_d,以及某个属性值发生并保持的开始时间 t_s 和结束时间 t_e,这里$[t_s,t_e)\subseteq[t_b,t_d)$,后者称为生命区间(lifespan);调查时间——感知现实目标而获取信息的时间,这是一个用户定义时间,记为 t_u;执行时间(或称事务时间,transaction time)——数据进入系统的时间。

我们不允许目标再生(reincarnation)。现实中固然存在着一些事物能在消亡后复生,但它们的属性已经不尽相同,往往还发生了质变,看作两个不同的目标也是合理的。即使它们的属性相同而适宜认作同一目标,也可以将那一段消亡当作一种特殊属性,其结果便得了一个连续

收稿日期:1995-10-12. 黄明智,男,36岁,博士,现从事地理信息系统研究。

的生命区间。事实上，这种情况下的消亡只是表像，前后生命区间之间必定存在着一种潜在的联系。

1.2 时态区间

设时间全集为 $T=[0,\text{now}]$，now 表示当前时间，[]是通常数学意义上的区间符号。

一般地，称 $I=[t_s,t_e)$ 为时态区间，$t_s,t_e\in T$。区间$[t_1,t_1)$和$[t_2,t_3)$相邻，$[t_1,t_2)$是$[t_2,t_3)$的前趋，$[t_2,t_3)$是$[t_1,t_2)$的后继。

这里我们规定时态区间的长度非 0，即存在着实数 $\varepsilon>0$，使 $t_e\geqslant t_s+\varepsilon$，这样，时间点(时刻)$t$与$[t,t+\varepsilon)$等同。这种规定是现实而合理的，至少我们可将 ε 理解为系统的时间分辨率。

now 是个重要的时间概念。$t_e\neq$now 的区间$[t_s,t_e)$是在调查时"能够"(当然未必是在实际上已经做到)确定地获知目标属性的区间，称为确定区间；对于区间$[t_s,\text{now}]$，因为在调查时 $\text{now}=t_u$，故$[t_s,t_u]$部分是确定的，而现在 $\text{now}>t_u$，$(t_u,\text{now}]$部分是在 $\text{now}=t_u$ 时"不能够"确定地获知的，它只是其前趋的一个延伸，即我们不确定地认为$[t_s,t_u]$中的目标属性在$(t_u,\text{now}]$中仍为真，$[t_s,\text{now}]$因此被称为不确定区间。

我们还要规定 $t_d\leqslant$now 及$\forall\ t_s,t_e(t_s\leqslant t_e\leqslant t_u)$，故$(t_b,t_d)\subseteq T$ 也是时态区间。

定义 $I_1<I_2$，若$\forall\ t_1,t_2(t_1\in I_1\wedge t_2\in I_2\rightarrow t_1<t_2)$。

2 时态赋值、时态元组和时态关系

2.1 时态赋值

我们先定义两个属性集：A 是目标的一个值域为 $\text{dom}(A)=\{a_i\}$ 的狭义属性集，其定义域为生命区间$[t_b,t_d)$中经过调查已经确知目标属性的时间区间，为$[t_b,t_d)$因而也是 T 的子集。即 A 是具有实际意义的目标属性集。E 为对应于狭义属性集 A(也可称 A 对应于 E)的广义属性集，其值域为 $\text{dom}(E)=\{e_i\}$。两者关系为：

$$E=A\cup\{\Theta,X,\Omega\},\qquad \Theta,X,\text{不属于 }A$$

这里，Θ、X、Ω 是 3 个广义属性值：Θ 表示 $t\notin[t_b,t_d)$，即目标消亡；X 表示 $t\in[t_b,t_d)$，但其狭义属性值未知；Ω 表示毫无所知。E 的定义域是时间全集 T。

A 和 E 的用户视图[6]和处理结果完全相同，差异只在于系统的表示和处理方法。以长度非 0 的时态区间和广义属性集 E 这两个概念为基础，本文对于时态赋值的定义和讨论远比文献[4,5]中的简单和实用。

函数

$$\{\zeta(t)=e_i\mid t\in I_i,i=1,\cdots,n\}\tag{1}$$

是 A 的调查时间 t_u 的一个时态赋值，其中 $I_1<I_2<\cdots<I_n$ 且 $I_1+I_2+\cdots+I_n=T$(即 I_i 与 I_{i+1}相邻)，$e_i\in\text{dom}(E)$。

我们称$[\![\zeta]\!]=[0,t_u]$为 ζ 的确定时态域，$\langle\!\langle\zeta\rangle\!\rangle=T-[\![\zeta]\!]=(t_u,\text{now}]$为 ζ 的不确定时态域。

例 1 某宗地利用类别的时态赋值 ζ_1 如表 1 所示。由于不知该宗地在[1940,1949)中何时生成，便将其定为 Ω 区间。

表 1 时态赋值 ζ_1

	区 间
Θ	[0,1940)
Ω	[1940,1949)
住宅	[1949,1956)
X	[1956,1970)
工厂	[1970,1978)
X	[1978,1983)
住宅	[1983,now)
调查时间	$t_u=1985$

不难理解，时态赋值的表达形式如：

$$[\Theta][\Omega]\{a_i\mid X\}[\Omega][\Theta]\tag{2}$$

其中[]中的项是不一定存在的;符号|隔开的左右两项任取其一;{}中的项是可重复的。

在时间轴的左端,一般是一个 Θ 区间。若 t_b 已知,则其必为 Θ 区间的右端点;否则,存在着一个包含 t_b 的 Ω 区间,或称 t_b 区间。而在时间轴的右端,也可能存在着包含 t_d 的 Ω 区间,或称 t_d 区间。若该区间的右端点 $t\leqslant$now,则还存在着 Θ 区间[t,now],否则 Θ 区间不存在。若右端区间[t,now]的值为 $a_i|X$,则 Ω 区间和 Θ 区间都不存在。

式(2)还说明,Ω 区间和 Θ 区间总是相邻成对的(如果均存在),后者在前者之外,而且最大对数为 2。这一点可用于逻辑检查。

为了简洁,不必要时可不列出 Ω 区间和 Θ 区间。

2.2 时态元组

一个时态数据库模式记为 $R=\{A_1,A_2,\cdots,A_n\}=A_s\cup A_t$,其中 A_s 和 A_t 分别是由静态型和时态型属性组成的子集。时态型属性又称为时变属性,其值随时间改变;静态型属性则不随时间改变,也称时不变属性。

R 上的时态元组 τ 是 R 上的这样一个函数,它使得对于每个属性 $A\in A_t$,τ 在 A 上的投影(或分量)$\tau\cdot A$ 是 A 的一个时态赋值。τ 与其中的每个属性都具有相同的生命区间[t_b,t_d),或 t_b 区间和 t_d 区间。我们记 $\tau(t)$是对应于时态元组 τ 和时间 $t\in T$ 的静态元组,它满足

$$\forall A_i\in R(\tau\cdot A_i)(t)) \tag{3}$$

显然,当 $t\notin[t_b,t_d)$时,$\tau(t)$是个空元组。

2.3 时态关系

设给定 $K\subseteq R$ 是 R 上的关键字,R 上有限个非空的时态元组的集合构成 R 上的一个时态关系 $r=\{\tau_i|i=1,2,\cdots,m,m<+\infty\}$,它满足:1)$K\subseteq A_s$,即 K 中只包含静态型属性;2)$\forall\ \tau_1,\tau_2\in r(\tau_1\neq\tau_2\rightarrow\exists\ A\in K(\tau_1\cdot A\neq\tau_2\cdot A))$,即不同元组的关键字不相同。我们将这两个条件称为时态范式(Temporal Normal Form)。我们称 $r(t)=\{\tau(t)|\tau\in r\}$是时态关系 r 在时间 $t\in T$ 的一个静态快照(snapshot)。

3 时态数据的比较与一致性

3.1 时态赋值的比较

设 ζ_1 和 ζ_2 是调查时间分别为 t_1 和 t_2 的时态赋值,将它们表示为可比较的形式:

$$<\zeta_1,\zeta_2>=\{<a_i,b_i,I_i>|i=1,\cdots,n\} \tag{4}$$

其中 $I_1<I_2\cdots<I_n$ 且 $I_1+I_2+\cdots+I_n=T$,$a_i,b_i\in\mathrm{dom}(E)$,而

$$a_i=\begin{cases}\zeta_1(I_i) & I_i\subseteq[\![\zeta_1]\!]\\ \#\zeta_1(I_i) & I_i\subseteq[\![\zeta_1]\!]\end{cases}\qquad b_i=\begin{cases}\zeta_2(I_i) & I_i\subseteq[\![\zeta_2]\!]\\ \#\zeta_2(I_i) & I_i\subseteq[\![\zeta_2]\!]\end{cases}$$

I_i 与 I_{i+1}相邻,但 $a_i\neq a_{i+1}$或 $b_i\neq b_{i+1}(i\leqslant n-1)$。

表 2　广义属性值的一致性比较及扩充结果选取

		延伸值#	未知属性 Ω	未知狭义属性 X	消亡 Θ	狭义属性 a_1
延伸值	#	前一区间值	前一区间值	前一区间值|X	Θ	a_1
未知属性	Ω		Ω	X	Θ	a_1
未知狭义属性	X			X	不一致	a_1
消亡	Θ		对　称		Θ	不一致
狭义属性	a_2					待比较?

注:除注明"不一致"或"待比较"之外的各项均为"一致",# 和 X 扩充的结果可选为"前一区间值"或 X,在系统中统一设定.

注意：这里将不确定区间[t_s,now]分开为[t_s,t_u]和[t_u,now]，前缀＃表示延伸值。若ζ_1和ζ_2符合表 2 定义的一致性，则称它们是一致的，记为$\zeta_1 \approx \zeta_2$；否则称它们不一致，$\zeta_1 \not\approx \zeta_2$。

从表 2 可见，时态赋值的一致性比较只是在确定区间内对狭义属性a_1、广义属性X和Θ进行比较，X和a_i被定义为一致的，但它们都与Θ不一致；延伸值＃和广义属性Ω被定义为与任何广义属性e_i都一致。

例 2　设例 1 中宗地利用类别的另一时态赋值ζ_2如表 3，则ζ_1、ζ_2可比较的形式如表 4 所示，其结果显示ζ_1、ζ_2是一致的。

表 3　时态赋值ζ_2

	区　间
Ω	[0,1974)
工厂	[1974,1983)
X	[1983,1987)
商用	[1987,1989)
Ω	[1989,now)
	t_u=1990

表 4　时态赋值的比较和扩充

ζ_1,ζ_2 的比较		$\zeta_3=\zeta_1 \perp \zeta_2$	
Θ,Ω	[0,1940)	→Θ	[0,1940)
Ω,Ω	[1940,1949)	→Ω	[1940,1949)
住宅,Ω	[1949,1956)	→住宅	[1949,1956)
X,Ω	[1956,1970)	→X	[1956,1970)
工厂,Ω	[1970,1974)		
工厂,工厂	[1974,1978)	→工厂	[1970,1983)
X,工厂	[1978,1983)		
住宅,X	[1983,1985)		
＃住宅,X	[1985,1987)	住宅	[1983,1987)
＃住宅,商用	[1987,1989)	→商用	[1987,1989)
＃住宅,Ω	[1989,now]	→Ω	
t_{u1}=1985,t_{u2}=1990		t_u=1990	

3.2　时态元组和时态关系的比较

对于时态数据库R上的时态元组τ_1和τ_2，若

$$\forall A \in A_s(\tau_1 \cdot A = \tau_2 \cdot A) \wedge \forall A \in A_t(\tau_1 \cdot A \approx \tau_2 \cdot A) \tag{5}$$

我们称它们一致，$\tau_1 \approx \tau_2$。对于R上的两个关系r_1和r_2，若

$$\forall (\tau_1 \in r_1 \wedge \tau_2 \in r_2)(\forall A \in K(\tau_1 \cdot A = \tau_2 \cdot A) \rightarrow \tau_1 \approx \tau_2) \tag{6}$$

我们称它们一致，$r_1 \approx r_2$。

4　时态数据的操作

4.1　时态赋值的扩充和更新

让我们再来看表 3。虽然X和a_i一致，延伸值＃和广义属性Ω与任何广义属性e_i都一致，但显然a_i、Θ比X、＃或Ω更确定。下面我们定义扩充操作。

设ζ_1和ζ_2是调查时间分别为t_1和t_2的时态赋值，若$\zeta_1 \approx \zeta_2$且至少在一个可比较的时间区中ζ_1与ζ_2的确定性不同，可将ζ_1扩充到$\zeta=\zeta_1 \perp \zeta_2$：

$$\left.\begin{array}{l}\text{调查时间 } t_u = \max\{t_1,t_2\};\\ \text{在可比较的每个时间区间中，}\zeta\text{ 的值按表 3 选取}\end{array}\right\} \tag{7}$$

显然，$\zeta_1 \perp \zeta_2 = \zeta_2 \perp \zeta_1$。

实际上，表 3 按照广义属性的确定性列出了不同的优先级：

$$\left.\begin{array}{l}a_i,\Theta\\ X \text{ 或 } \#\text{（系统统一规定）}\\ \Omega\end{array}\right\} \tag{8}$$

必须注意：如果前一区间的扩充结果与被延伸值不同，则该区间不能延伸该值，即将＃的优先级降为Ω，说明见例 3。

例 3　在例 2 中$\zeta_1 \approx \zeta_2$，但ζ_1与ζ_2的确定程度不同，可将ζ_1扩充为$\zeta_3=\zeta_1 \perp \zeta_2$，结果仍见

表 4。其中,区间[1985,1987)的扩充结果是 ζ_1 对其前趋的延伸值"住宅";但对于区间[1989,now]来说,因为其前趋[1987,1989)的扩充结果已为"商用",故而不可以延伸"住宅",扩充结果为 Ω。

不论 ζ_1 和 ζ_2 是否一致,可用 ζ_1 更新 ζ_2(一般地应有 $t_1 \geqslant t_2$)为 $\zeta=\zeta_1 // \zeta_2$:

$$\left.\begin{array}{l}\text{调查时间 } t_u=\max\{t_1,t_2\};\\ \text{在可比较的每个时间区间中,若 } \zeta_1 \approx \zeta_1\text{,则更新结果与表 3 所列的扩充结果相同;}\\ \text{否则其更新结果为 } \zeta_1 \text{ 的值。}\end{array}\right\} \quad (9)$$

显然,$\zeta_1 \approx \zeta_2$ 时 // 操作等同于 $\perp$ 操作;$\zeta_1 \not\approx \zeta_2$ 时 $\zeta_1 // \zeta_2 \neq \zeta_2 // \zeta_1$。

例 4　设例 1 中宗地利用类别的另一时态赋值 ζ_4 见表 5,经比较知 $\zeta_1 \not\approx \zeta_2$,可用 ζ_4 更新 ζ_1 得 $\zeta_5=\zeta_4 // \zeta_1$,见表 6。注意,该例中省去了两个 Θ 区间[0,1940)和[1987,now]。

表 5　时态赋值 ζ_4

	区　间
Ω	[0,1956)
住宅	[1956,1975)
X	[1975,1984)
商用	[1984,1987)
	t_u=1988

表 6　时态赋值的更新 $\zeta_5=\zeta_4 // \zeta_1$

	区　间
Ω	[1940,1949)
住宅	[1949,1975)
工厂	[1975,1978)
X	[1978,1983)
住宅	[1983,1984)
商用	[1984,1987)
	t_u=1988

4.2　时态元组的扩充和更新

设在 R 上有 $\tau_1 \approx \tau_2$,我们定义元组的扩充 $\tau=\tau_1 \perp \tau_2$。τ 的分量值为:

$$\tau \cdot A=\begin{cases}\tau_1 \cdot A=\tau_2 \cdot A, A \in A_s\\ \tau_1 \cdot A \perp \tau_2 \cdot A, A \in A_t\end{cases} \quad (10)$$

如果 τ_1 和 τ_2 具有相同的关键字,即 $\forall A \in K(\tau_1 \cdot A=\tau_2 \cdot A)$,则不论 τ_1 和 τ_2 是否一致,我们定义元组的更新 $\tau=\tau_1 // \tau_2$。τ 的分量值为:

$$\tau \cdot A=\begin{cases}\tau_1 \cdot A=\tau_2 \cdot A, A \in K\\ \tau_1 \cdot A \quad\quad , A \in A_s-K\\ \tau_1 \cdot A // \tau_2 \cdot A, A \in A_t\end{cases} \quad (11)$$

4.3　时态关系操作

设在 R 上有 $r_1 \approx r_2$,我们定义扩充操作(即传统关系中并操作的扩展):

$$\begin{aligned}r=r_1 \perp r_2=\{\tau \mid &(\exists\, \tau_1 \in r_1, \exists\, \tau_2 \in r_2(\forall\, A \in K(\tau_1 \cdot A=\tau_2 \cdot A) \wedge \tau=\tau_1 \perp \tau_2))\\ &\vee (\tau \in r_1 \wedge (\forall\, \tau_2 \in r_2(\exists\, A \in K(\tau \cdot A \neq \tau_2 \cdot A))))\\ &\vee (\tau \in r_2 \wedge (\forall\, \tau_1 \in r_1(\exists\, A \in K(\tau \cdot A \neq \tau_1 \cdot A))))\}\end{aligned} \quad (12)$$

不论 r_1 与 r_2 是否一致,我们定义以 r_1 更新 r_2 的操作 $r=r_1 // r_2$。表达式的形式与(12)同,只是将符号 $\perp$ 改为 //。

至此,我们定义了时态属性、时态元组和时态关系的更新操作。

5　结　语

1)以往的时态模型研究中对于不确定时间值"now"没有很好处理,文献[5]也只是将[t_s,

now]看作一个向当前时间延伸的区间。引进了调查时间 t_u，将[t_i,now]分割为确定的[t_i,t_u]和不确定延伸的(t_u,now]两部分，显然是一种更为准确、完善的处理方法。以此为基础才得以定义数据更新的关系运算。

2)文献[4]对于时态元素和时态赋值提出了严格的数学条件，并且要求一个元组中各属性的时态域必须相同。根据实用的时态数据库中时间的结构特点，作者采用了广义属性集和区间赋值的规范化表达方式，从而简化和降低了上述要求，提高了适应性。同时，作者将时态型属性和静态型属性区分开来，并且强调了时态目标的生命区间，使属性、元组及关系的定义和操作更准确、更充分。

3)经过实验分析，本文定义的更新操作是严格、合理、有效的。然而，N1NF 模型毕竟比 1NF 模型更复杂些，虽然已经开发出不少实验型 DBMS[1,8]，要推出像 dBASE 那样的商品化系统则还须待以时日。

参 考 文 献

1 田沧海，林钧海. 支持多介质数据库的 NF2 方法. 计算机科学，1992，9(1)

2 Levene M, Loizou G. The Nested Relation Type Model: An Application of Domain Theory to Databases. The Computer Journal, 1990,33(1):19～30

3 Mark A R, Henery F K. Extended Algebra and Calculus for Nested Relational Databases. ACM DS, 1988,13(4):389～417

4 Gadia S K, Yeung C. Inadequacy of Interval Timestamps in Temporal Databases. Information Sciences, 1991,54(12):1～22

5 张师超. 时态关系代数与元组演算的等价性. 计算机学报，1993，16(12)

6 萨师煊，王 珊. 数据库系统概论(第二版). 北京：高等教育出版社，1991.

7 张祖勋，黄明智. 时态 GIS 的概念、功能和应用. 测绘通报，1995(2)

8 柳诚飞，龚正良. 一个基于扩充的 NF2 模型的 DBMS. 计算机学报，1992，15(12)

A Renewal Operation of N1NF Temporal Databases

Huang Mingzhi Zhang Zuxun

(Dept. of Surveying Engineering, Hehai University, 1 Xikang Road, Nanjing, China, 210098)

Abstract In this paper a user-defined time named "survey time" is introduced, which makes the manipulation of the indeterminal time "now" more accurate and proper. And an relational algebra-based renewal operation is also defined.

Key words temporal database; non first a normal form (N1NF); data renewal

(上接 138 页)

GIS are discussed in this paper. And accuracy indicators for estimating positional error of area primitives and area of uncertainty region are also given. These are of some value to both further researches and industrial production.

Key words area primitive; positional error; accuracy indicator