50—54 时态,数据库 数据库 时态关系代数 10

计算机科学 Vol.20 No.1 1993

一个时态关系代数

张师超

TP311.13

(国家智能计算机研究开发中心 广西师范大学数学系 541004桂林)

摘 要

In the current temporal relational models, it leads to inadequate of the operations that the temporal relational operations act temporal data as well as static data except relational union. This paper sets up a N_1NF temporal relational algebra. Its operations are sufficient for the needs of varied applications.

1.引言

随着越来越多的私人和政府部门等需要保存和查看过去的记录(或数据),时态数据库已成为一个活跃的研究领域。据统计,近十多年来,已有400余篇有关时间信息处理的文章发表。这些研究主要集中在:(1)时态数据模型;(2)时态询问语言;(3)存贮结构和实现技术。

由于时间具有一些特有的性质,它不象一般数据那样易于在数据库中模拟。所以,许多研究人员认为,有必要从零开始建立时态数据库的理论与基础[2]。L. E. Mckenzie和R. T. Snodgrass[1]用26种评价标准测试了12个典型的时态数据库后,其结果与上面一致。

从现有的时态关系模型看,它仅仅是传统的关系模型的一个简单扩展。除关系并外,其它的关系操作仍不能体现时态的特性,这导致时态操作的不充分性。本文建立一个

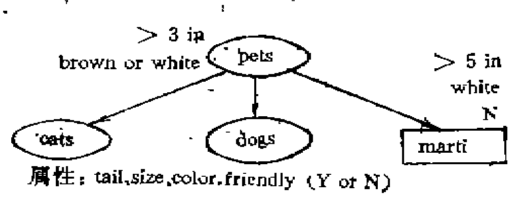

图2 动物模式

:N〉

D_2= 〈Year:1988, Insured:Y, Color: white, Owner:john〉

D_3= 〈Year:1990, Insured:Y, Color:yellow, Owner:john〉

d-join(D_1, D_2) = 〈Tail:8 in, Color: white, Friendly:N, Year:1988, Insured:Y, Owner: john〉

d-join(D_1, D_3) = ε

2. 有一个项是实例:分解为

d-join(($C_1 - D_{1,1} - D_{1,2} - \cdots - D_{1,k}$), D_2) = d-join(C_1, D_2) — [d-join($D_{1,1}$, D_2) + ⋯ + d-join($D_{1,2}$, D_2) + ⋯ + d-join($D_{1,k}$, D_2)]

3. 两个项均为具有缺省值的类,将项当成实例对待。

4. 两个项均为概念项:

d-join(($C_1 - D_{1,1} - D_{1,2} - \cdots - D_{1,k}$), ($C_2 - D_{2,1} - D_{2,2} - \cdots - D_{2,k}$)) = d-join($C_1$, C_2) — (d-join(C_1, $D_{2,1}$) + ⋯ + d-join(C_1, $D_{2,k}$)) — (d-join(C_2, $D_{1,1}$) + ⋯ + d-join(C_2, $D_{1,k}$))

(参考文献略)

[冯铃摘译自"Proceedings of the 17th Intl. Conf. on VLDB",冯玉才校]

NINF时态关系代数，其操作是充分的。

2．时态模型

为简化问题的描述，我们假定涉及到的时间全集为$C=[0, NOW]=\{0, 1, 2, \cdots, NOW\}$，并且C满足线性序$\leqslant$，NOW表示当前时间。

定义2.1 对于一个集合$I\subseteq C$，如果$\forall t_1, t_2, t_3\ (t_1\leqslant t_2\leqslant t_3\wedge t_1\in I\wedge t_3\in I\longrightarrow t_2\in I)$，则称I为C的一个区间。

定义2.2 设P是C上的全部区间组成的集合，Φ和C分别为P的最小和最大元素。对于P中的任意的n个元素$A_1, A_2, \cdots, A_n\in P$，且$n<+\infty$，$A_1, A_2, \cdots, A_n$的并称为C上的一个时态元素(temporal element)。

显然，$\forall A\in P$，A是一个时态元素。一个时刻$t\in C$可以认为是一个时态元素，记为$[t, t]$。

我们用μ, ν, μ_1, ν_1, …来表示时态元素，时态元素μ和ν的并和交分别表示为$\mu+\nu$和$\mu*\nu$，且μ的补$(C-\mu)$表示为$-\mu$。它们的结果显然也是时态元素。

性质2.1 设C上的所有时态元素组成的集合用TE表示，且Φ和C分别是TE的最小和最大元素，则TE在+、*、－运算下是一个布尔代数。

对于时态元素μ和ν，我们说$\mu\leqslant\nu$（或$\mu<\nu$），如果所有的$t_1\in\mu$和$t_2\in\nu$有$t_1\leqslant t_2$（或$t_1<t_2$）。我们允许在时态元素间作集合的比较$\subseteq$和=操作。

定义2.3 设μ和ν是时态元素。i) μ和ν的差为$\mu-\nu=\{t\mid t\in\mu\wedge t\notin\nu\}$，ii) μ中的第一个和最后一个时刻分别记为first(μ)和last(μ)。定义$[\mu,\nu]=[\text{first}(\mu),\text{last}(\mu)]$，iii) 设$B_1=\{A_1\mid A_1$是$\mu$上的一个区间，且first$(\mu)\in A_1\}$，$B_2=\{A_2\mid A_2$是$\mu$上的一个区间，且last$(\mu)\in A_2\}$。用$f_s(\mu)$和$l_s(\mu)$分别表示$B_1$和$B_2$中的最大元素，即$f_s(\mu)$和$l_s(\mu)$分别是$\mu$的第一个和最后一个最大区间。

定义2.4 设$\mu=A_1+A_2+\cdots+A_n$，如果$\forall i, j\ (1\leqslant i<j\leqslant n\longrightarrow A_i<A_j)$，则称$\mu$为规范的时态元素。

我们假设本文涉及的时态元素均为规范的时态元素。

例2.1 设$\mu=[0, 5]+[8, 8]$，$\nu=[3, 10]$是两个时态元素。则有：

first$(\mu)=0$，last$(\mu)=8$，first$(\nu)=3$，last$(\nu)=10$，$[\mu, \nu]=[\text{first}(\mu), \text{last}(\nu)]=[0, 10]$，$f_s(\mu)=[0, 5]$，$l_s(\mu)=[8, 8]$，$f_s(\nu)=l_s(\nu)=[3, 10]$。

为了反映事物的变化性质，我们引进时态赋值的概念。

定义2.5 设A是一个属性，它的值域记为dom(A)。A的一个时态赋值ξ是从μ到dom(A)的一个函数，它使得每个$t\in\mu$，$\xi(t)\in\text{dom}(A)$。这里，$\mu\in C$是一个时态域。

下面，我们用$\overline{\xi}$表示属性A的时态赋值ξ的时态域，用$|\xi|$表示ξ的范围$\{\xi(t)\mid t\in\overline{\xi}\}\subseteq\text{dom}(A)$。如果$\nu$是一个时态元素，我们用$\xi\restriction\nu$表示$\xi$在$\nu*\overline{\xi}$上的限制。显然，限制$\xi\restriction\nu$也是A的一个时态赋值。

定义2.6 设ξ_1和ξ_2是属性A的时态赋值。

i) 如果ξ_1和ξ_2一致，即$\forall t\in\overline{\xi_1}*\overline{\xi_2}\ (\xi_1\restriction[t, t]=\xi_2\restriction[t, t])$，那么，$\xi_1\cup\xi_2$是$\xi_1$和$\xi_2$到时态域$\overline{\xi_1}+\overline{\xi_2}$上的一个一般扩充，这个扩充也是A的时态赋值。

ii) 设$\mu=\{t\mid t\in\overline{\xi_1}*\overline{\xi_2}\wedge\xi_1(t)=\xi_2(t)\}$。我们定义$\xi_1\cap\xi_2$为$\xi_1\restriction\mu$（或$\xi_2\restriction\mu$），且$\xi_1-\xi_2$为$\xi_1\restriction(\overline{\xi_1}-\mu)$。

为了方便，我们将对属性A的一个赋值ξ表示为$\xi=\langle\eta_1, \eta_2, \cdots, \eta_l\rangle=\langle\langle\nu_1, a_1\rangle, \cdots, \langle\nu_l, a_l\rangle\rangle$，这里，$\eta_i$是$\xi$的元素，$\nu_i$是时态元素，$a_i\in\text{dom}(A)$，$i=1, 2, \cdots, l$，且$\xi(t)=a_i$，如果$t\in\nu_i$，$1\leqslant i\leqslant l$，以及$a_i\neq a_j$如果$i\neq j$。

当然，上面的$\eta_i=\langle\nu_i, a_i\rangle$也是A的一个赋值，$i=1, 2, \cdots, l$，所以，我们可以认为$\xi$是A的几个象$\eta_i$这种简单赋值的集合。$\xi$也可以只含有一个元素。

定义2.7 假定ξ_1和ξ_2分别是θ可比较属

性A和B的赋值，定义$\xi_1 \ \theta \ \xi_2=\{t \mid t\in\overline{\xi_1}*\overline{\xi_2} \wedge \xi_1(t) \ \theta \ \xi_2(t)\}$。

现在，我们来定义时态数据库的元组和关系。

定义2.8 一个数据库模式R上的元组τ是$\dot{R}$上的这样一个函数，它使得对于R的每个属性A，$\tau(A)$是A的一个时态赋值。

我们假定在后面讨论的时态数据库中没有空值元组。

定义2.9 设R是一个数据库模式，给定$K\subseteq R$是R上的键。那么R上的一个关系r是R上有限个非空元组的集合，它满足：

(1) 对于r中的任意元组τ及$A\in K$，$|\tau(A)|$是唯一的，即$|\tau(A)|$中仅有一个值；

(2) 对于r中任意元组τ_1和τ_2，$\forall A\in K$ ($|\tau_1(A)|=|\tau_2(A)|$ 当且仅当$\tau_1=\tau_2$。

通常，定义2.9中r应满足的两个条件称为时态范式(temporal normal form)。

性质2.2 设时态数据库模式R上的元组集r和键集$K\subseteq R$被给定，那么，对于每个$t\in C$，$r\upharpoonright[t,t]=\{\tau\upharpoonright[t,t] \mid \tau\in r\}$是r在时刻t的快照（或快照数据库）。

3．关系代数

为了达到安全的时态关系操作能力，我们需要对已有的并、差、笛卡尔积、投影、和选择这几个基本的操作上增加两个操作符。它们是同键值归并($\oplus$)和元组分取($\odot$)两个操作。

(1) 同键值归并$\oplus$操作：对于时态数据库模式R上的两个关系r'和r''。设$K=\{K_1, K_2, \cdots, K_n\}\subseteq R$为R上的键集合，$A=R-K=\{A_1, A_2, \cdots, A_m\}$为R上的非键集合。将$r_1'$和$r''$合为关系r时，如果有$\tau'\in r'$和$\tau''\in r''$在K上有，对于$\forall K'\in K$，$|\tau'(K')|=|\tau''(K')|$，则$\tau'$和$\tau''$合并为r中的元组$\tau$时需要$\oplus$操作。经$\oplus$操作后，$\tau$在K和A上的投影为：

i) $\tau(K)=\{\langle\overline{\tau'(K_1)}+\overline{\tau''(K_1)}, |\tau'(K_1)|\rangle, \cdots, \langle\overline{\tau'(K_n)}+\overline{\tau''(K_n)}, |\tau'(K_n)|\rangle\}$；

ii) 在$\tau(A)$中，对于$1\leqslant i\leqslant m$，设$F=\tau'(A_i)=\{\langle\mu_1, a_1\rangle, \cdots, \langle\mu_l, a_l\rangle\}$，$H=\tau''(A_i)=\{\langle\nu_1, b_1\rangle, \cdots, \langle\nu_k, b_k\rangle\}$。则，

a) 对于任意的$1\leqslant x\leqslant l$，$a_x\notin|\tau''(A_i)|$，将$\langle\mu_x, a_x\rangle$增加到集合G中，并将$\langle\mu_x, a_x\rangle$从F中去掉；

b) 对于任意的$1\leqslant y\leqslant k$，$b_y\notin|\tau'(A_i)|$；将$\langle\nu_y, b_y\rangle$增加到G中并将$\langle\nu_y, b_y\rangle$从H中去掉；

c) 对于任意的$1\leqslant z\leqslant l$，若存在一个$j(1\leqslant j\leqslant k)$使得$a_z=b_j$，则将$\langle\mu_z+\nu_j, a_z\rangle$增加到G中，并分别将$\langle\mu_z, a_z\rangle$和$\langle\nu_j, b_j\rangle$从F和H中删除；

d) 若存在$\langle\mu', c\rangle$和$\langle\mu'', d\rangle\in G$，且$last(\mu')=last(\mu'')=NOW$。根据上述操作，显然$c\neq d$。如果$first(l_s(\mu'))<first(l_s(\mu''))$，则用$[first(l_s(\mu')), first(l_s(\mu''))-1]$代替$\mu'$的$l_s(\mu')$；如果$first(l_s(\mu'))>first(l_s(\mu''))$，则用$[first(l_s(\mu'')), first(l_s(\mu'))-1]$代替$\mu''$的$l_s(\mu'')$；如果$first(l_s(\mu'))=first(l_s(\mu''))$，则整个合并失败。这种情况是因$r'$和$r''$不一致产生的，应予以避免。

e) $\tau(A_i)=G$。

我们记τ_1和τ_2的$\oplus$操作为$\tau_1\oplus\tau_2$。

(2) 元组分取$\odot$操作：元组分取$\odot$操作的假设与$\oplus$操作相同，而$\odot$操作是取出元组$\tau'\in r'$和$\tau''\in r''$的一致部分。下面，我们来定义$\tau=\tau'\odot\tau''$。

i) $\overline{\tau(K_i)}=\overline{\tau'(K_i)}*\overline{\tau''(K_i)}$，$|\tau(K_i)|=|\tau'(K_i)|$；

ii) $\overline{\tau(A_j)}=\{t \mid t\in\overline{\tau'(A_j)}*\overline{\tau''(A_j)} \wedge (\tau'(A_j)\upharpoonright[t,t]=\tau''(A_j)\upharpoonright[t,t])\}$，$\tau(A_j)=\tau'(A_j)\upharpoonright\overline{\tau(A_j)}$。

注意，上面及后面假设的关系在不做特殊说明的情况下，关系都是满足时态范式的。

4．基本时态关系代数操作的扩展

为了使得数据库系统提供的操作能处理数据的时态，我们需要扩展传统的基本关系代数操作。

1）扩展并（$\cup^e$）

设r_1和r_2是时态数据库模式R上的两个关系，且r_1和r_2是一致的。$K\subseteq R$是R的键集，$A=R-K$是R的非键集。那么，

$r=r_1\cup^e r_2=\{\tau\mid(\exists\tau_1\in r_1,\ \exists\tau_2\in r_2\ (\forall K'\in K(\mid\tau_1(K')\mid=\mid\tau_2(K')\mid\wedge\tau=\tau_1\oplus\tau_2)))\vee(\tau\in r_1\ \wedge\ (\forall\tau'\in r_2\ (\exists K'\in K(\mid\tau(K')\mid\neq\mid\tau'(K')\mid))))\vee(\tau\in r_2\wedge(\forall\tau'\in r_1(\exists K'\in K(\mid\tau(K')\mid\neq\mid\tau'(K')\mid)))))\}$。

2）扩展差（$-^e$）

扩展差的假设与扩展并相同。那么，

$r=r_1-^e r_2=\{\tau\mid(\exists\tau_1\in r_1,\ \exists\tau_2\in r_2(\forall K'\in K(\mid\tau_1(K')\mid=\mid\tau_2(K')\mid\wedge(\tau_1\ (R)\ \textcircled{I}\ \tau_2\ (R)\neq\Phi\wedge\tau=(\tau_1\restriction\ (\overline{\tau_1}-\overline{\tau_1\textcircled{I}\tau_2}))\restriction(\overline{\tau_1}\ \ominus\ l_s(\overline{\tau_2})))\vee(\tau_1(R)\textcircled{I}\tau_2(R)=\Phi\wedge\tau=\tau_1\restriction(\overline{\tau_1}\ominus l_s(\overline{\tau_2})))))\vee(\tau\in r_1\wedge(\forall\tau'\in r_2(\exists K'\in K(\mid\tau(K')\mid\neq\mid\tau'(K')\mid)))))\}$。

注意，上面式子中的"－"号的的意义与第2节中时态元素差相同。现来定义"$\ominus$"操作如下：

定义4.1 设μ_1和μ_2为两个时态元素，我们定义$\ominus$操作如下：

i）$NOW\in l_s(\mu_1)$，$NOW\in l_s(\mu_2)$且$l_s(\mu_1)\supset l_s(\mu_2)$，则$\mu_1\ominus\mu_2=(\mu_1-l_s(\mu_1))+(l_s(\mu_1)-l_s(\mu_2))$；

ii）其它情况下，$\mu_1\ominus\mu_2=\mu_1$。

3）扩展交（$\cap^e$）

扩展交的假设与扩展并相同。那么，

$r=r_1\cap^e r_2=r_1-^e(r_1-^e r_2)$.

4）扩展笛卡尔积（Etimes）

设r_1和r_2分别为时态数据库模式R和S上的关系。那么$r=r_1$ Etimes r_2是R S上的一个关系，并且r的每个元组$\tau=(a_1,\ a_2,\ \cdots,\ a_n,\ b_{n+1},\ b_{n+2},\ \cdots,\ b_{n+m})$中，$(a_1,\ a_2,\ \cdots,\ a_n)$是$r_1$的元组，$(b_{n+1},\ b_{n+2},\cdots,\ b_{n+m})$是$r_2$的元组。我们将r的元组简记为$\tau=(\tau_1,\ \tau_2)$，$\tau_1\in r_1$，$\tau_2\in r_2$。设$K_1\in R$是R的键集，$K_2\in S$是S的键集，则$K_1K_2$是RS的键集。则定义：

$r=r_1$ Etimes $r_2=\{\tau\mid\tau=(\tau_1,\ \tau_2)\wedge\tau_1\in r_1\wedge\tau_2\in r_2\}$.

5）扩展自然连接（$\bowtie^e$）

设r_1和r_2分别为时态数据库模式R_1和R_2上的关系，要对r_1和r_2做连接操作，必须有$R_1\cap R_2\neq\Phi$。我们设$X=R_1\cap R_2$，$A=R_1-X$，$B=R_2-X$，那么，$R=(A,\ X,\ B)$就是连接结果的关系r的时态数据库模式。r应满足：

$r=r_1\bowtie^e r_2=\{\tau\mid(\forall\tau_1\in r_1\wedge\exists\tau_2\in r_2\ (\overline{\tau_1}*\overline{\tau_2}\neq\Phi\wedge\tau(A)\restriction\mu\wedge\tau(B)=\tau_2(B)\restriction\mu\wedge\tau(X)=\tau_1(X)\restriction\mu\wedge\tau\neq\Phi\wedge\mu=\{t\mid t\in\overline{\tau_1}*\overline{\tau_2}\wedge\ (\tau_1(X)\restriction[t,\ t]=\tau_2(X)\restriction[t,\ t])\})))\}$.

我们设K_1和K_2分别是R_1和R_2的键集，则R的键集为$K=K_1\cup K_2$。

6）扩展投影（$\prod^e$）

扩展投影是使用扩展并$\cup^e$操作将传统投影结果合并成一个关系。设r_1是时态数据库模式R_1上的关系，$Y\subseteq R_1$，则在Y上的扩展投影记为$r=\prod^e_Y(r_1)$，r是R＝Y上的关系。那么，

$$r=\prod\nolimits^e_Y(r_1)=\bigcup^e_{\tau\in\prod_Y(r_1)}(\tau).$$

7）扩展选择（σ^e）

设r_1是时态数据库模式R上的一个关系。传统的选择操作(σ)是：$r_2=\sigma_F(r_1)$，其中F为谓词，它表达一个限制。扩展选择定义为：

$r=\sigma^e_{F\wedge T}(r_1)$.

F与其在传统选择操作中的意义相同，T是一个时态表达式，它是一个时间限制。

8）时态关系代数操作的封闭性

至此，我们已给出了时态关系代数操作的定义。现在，我们来证明这些操作对时态范式是封闭的。

定理4.1 扩展并、交、差、笛卡尔积、自然连接、投影、和选择操作对时态范式是封闭的。

证明：我们只证明扩展并对时态范式是封闭的，而其它操作的封闭性证明与此类似。

我们要证明扩展并对时态范式是封闭的，即要证明，对于任意的R上的两个关系r_1和r_2，$r=r_1\cup{}^t r_2$满足时态范式。(证明过程略)

5.结束语

文[3]指出了现有时态数据库中关系差操作是不充分的。这主要是因为现有的时态关系代数在将传统的关系代数扩展成能处理时间的时候，对"NOW"这个代表不定的（或变化的）时间值考虑不够，仅仅在扩展并操作中考虑了两个关系并操作后对"NOW"的处理，而象扩展差和扩展交中被忽略，因此，它导致关系操作的不充分性。

现在，在时态推理中已将"NOW"单独进行研究[4]，并且其处理是复杂的。当然，"NOW"在时态数据库也应被仔细地研究和探讨。作为一个尝试，笔者除了在关系并中考虑了"NOW"的作用，也在差中考虑了"NOW"的作用，并定义关系交为"$r=r_1-(r_1-r_2)$"，在关系交中间接地考虑到了"NOW"。这样，本文给出的时态关系代数克服了文[3]提及现有时态关系模型的缺点。

由于应用的需要，及光盘技术的进步使得存贮大量数据成为可能，所以，时态数据库方面的研究势在必然地成为一个具有广阔前景的课题。不过，时态数据库的研究在国内尚刚刚起步，笔者藉以本文抛砖引玉。

参考文献

[1] L. E. Mckenzie & R. T. Snodgrass, Evalution of Relational Algebras Incorporating the Time Dimension in Database, ACM Computing Surveys, Vol23, N0.4(1991)

[2] S. Gadia & C. Yeung, Inadequacy of Interval timestamps in Temporal Databases, Inform. Sciences Vol, 54,N0.1—2(1991)

[3] 张师超，时态数据库述评，计算机科学，Vol.19, No.3(1992)

[4] E.Hajniciz,Role of the present in temporal representation in artificial intelligence, Int.J. Man-Machine Studies, 32(1990)

（上接59页）

$$\leqslant \min_{r+s=j+1} [\lambda_r(A)+\lambda_s(B)]\ i=1,\cdots,n, \quad (9)$$

定理4 实对称矩阵A的最大特征值λ_1满足

$$\lambda_1(A)\leqslant \| A+\|A\|E\| - \|A\| \quad (10)$$

式中E是单位矩阵，若上式中$\|*\|$选取得当，(10)式的等号(接近)成立。用这个定理可以实现最小误差估算。

证明：略

以上只估算了对称矩阵的特征值，而对于一般的实矩阵的特征值的实部是否为负，可以通过定理5来判别。

定理5 A是任意的实矩阵。若对任意非零实向量x，有$x^TAx<0$，(当$x=0$时，$x^TAx=0$)，则矩阵A的特征值的实部为负。

证明：略。

因此，要使(6)式中的实矩阵P的特征值的实部是负数，只需证明对任意非零实向量x，有$x^TPx<0$（当$x=0$时，$x^TPx=0$），而$x^TPx=x^T[(P^T+P)/2]x$。

若实对称矩阵$(P^T+P)/2$的特征值小于0，则$x^TPx=x^T[(P^T+P)/2]x<0$。

求$(P^T+P)/2$的最大特征值，可以通过定理4来估算。这样对任意实矩阵的特征值的实部是否小于0的问题转化为对实对称矩阵的特征值的估算。这种方法用来检验系统的平衡点是否稳定简便而有效，而且很容易实现计算机模拟。

主要参考文献

[1] A.N.Michel et al., "Qualitative analysis of neural network," IEEE Trans. CAS, Vol36, No.2, P229-243, 1989

[2] 王联、王慕秋，"非线性常微分定性分析"，哈尔滨工业大学出版社，1987

[3] R.A.Hone和C.R.John著，杨奇译"矩阵分析"，天津大学出版社，1988