第5卷第9期　　软　件　学　报　　Vol.5, No.9
1994年9月　　JOURNAL OF SOFTWARE　　Sept. 1994

一个模糊时态关系代数*

张师超　　罗旭东

(国家智能计算机研究开发中心,北京 100080)　(北京轻工业学院自动化工程系,北京 100037)
(广西师范大学数学系,桂林 541001)

摘要　时间无处不在的特性要求数据库管理系统提供模拟现实世界中与时间有关的数据的设施.本文建立了一个模糊时态关系代数,它避免了本文中提到的现有时态数据库的一个弱点.

关键词　时态数据库,历史数据库,关系代数,模糊数据库.

事件在一特定的时间发生,事件与事件的关系也在某个特定的时间存在.因此,时间无处不存在,它是哲学、认识科学、物理学、计算机科学等必须要解决的主要问题.

对时态数据库的研究起于70年代末期.由于在许多应用,象预测与决策、会计等要求数据库管理系统提供支持现实世界中与时间有关的事件的设施.时态数据库的研究引起了广泛的兴趣,到现在已有近百个研究小组,约有400多篇这方面的文章发表,建立了近20种模型.由于关系模型具有广泛的应用基础,这些研究大多数是基于关系环境的.它们集中在概念级的时态形式语义,开发类似传统数据库的时态数据库模型,和时态查询语言的设计3个方面.对数据库增加时间的方式有时间点、区间、时态元素三种.

最近,L. Mckenzie和R. Snodgrass采用26个评价标准检测了12个典型的时态数据库模型.结果表明,没有一个模型满足全部的15个基本条件.另一方面,现实世界中的数据常常以模糊的形式存在,现在还没有看见模糊时态数据库方面的文章发表.此外,现有时态系统均是待试验的模型,完整性、规范化以及查询优化方面的工作很少.特别的,现有时态关系操作对"*NOW*"的处理不一致,这导致一些操作结果有误.

文献[1]建立了一个时态关系代数,本文建立与之相应的模糊时态关系代数.

1　模糊数据表示

表示模糊数据的方法已有许多种.本文采用单个数值表示数据的模糊方面,其方法在下面逐步介绍.

定义1. 设X是一个模糊数据,我们用二元组形式表示:$x=(c,b)$,这里,c为x的内容,b为x中所有模糊词的量化值且$b\in[0,1]$.

* 本文1992—07—14收到,1993—10—18定稿

本文受清华大学智能技术与系统实验室资助.作者张师超,32岁,副教授,主要研究领域为人工智能与时态数据库.罗旭东,31岁,讲师,主要研究领域为人工智能,程序变换,CAD.

本文通讯联系人:张师超,桂林541001,广西师范大学数学系

例如，x 表示"下大雪"用二元组表示为：$x=$（下雪，0.9）.

用二元组形式表示模糊数据主要方便于后面的描述. 其它表示方法的处理是类似的.

定义 2. 设 $x=(c_1,b_1)$，$y=(c_2,b_2)$ 为 2 个模糊数据. 记 x 和 y 的距离为：

$$\| x-y \| = \| c_1-c_2 \| + \| b_1-b_2 \|$$

其中 c_1 和 c_2 的距离按模糊数学中求距离的方式求值.

一个模糊数据 $x=(c,b)$ 与时间的关系实际上是的模糊值与时间的关系. 当 $b=0$ 时表示 x 不存在. 如果我们用 t 表示时间，那 x 和 t 的关系可用函数 $f_{(b,t)}$ 表示，或者 $b=g_{(t)}$. 通常，在 x 存在的时期内，$g_{(t)}$ 是常数函数，或者阶梯函数，或者是其它函数 3 种之一. 在数据库应用中，通常仅有前 2 种. 若确实碰到第 3 类函数，使用近似的阶梯函数代替，这可使用逼近技术让误差量减少.

2 时态模型

为便于问题描述，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 3. 对于一个集合 $I\subseteq C$，如果 $\forall\ t_1,t_2,t_3(t_1\leqslant t_2\leqslant t_3 \wedge t_1\in I \wedge t_3\in I \rightarrow t_2\in I)$，则称 I 为 C 的一个区间.

区间在并和补运算下是不封闭的，所以，它不能充分模拟包括"或"和"非"的数据及自然语言询问. 下面定义的时态元素可以克服区间的这一缺点.

定义 4. 设 P 为 C 上全部区间组成的集合，Φ 和 C 分为 P 的最小和最大的元素. 对于 P 中的任意 n 个元素 $I_1,I_2,\cdots,I_n\in P$，且 $n<+\infty$，u 为 $I_1,I_2,\cdots,I_n$ 的并，则称 u 为 C 上的一个时态元素，并记为 $u=I_1+I_2+\cdots+I_n$.

根据上面的定义，u 的表达是不唯一的，或者是不规范的. 因为可能有 $I_i\cap I_j\neq\Phi$，其中 $1\leqslant i,j\leqslant n$，因此，有必要定义 u 的唯一表达方式.

设 I_1 和 I_2 是两个区间，我们说 $I_1<I_2$，如果所有的 $t_1\in I_1$ 和 $t_2\in I_2$，有 $t_1<t_2$.

定义 5. 设 $u=I_1+I_2+\cdots+I_n$ 是 C 上的一个时态元素，若 u 是规范的，u 应该满足：

① $I_i\cap I_j\neq\Phi$，对于 $i\neq j$，$1\leqslant i,j\leqslant n$.

② $I_i<I_j$，如果 $i<j$，$1\leqslant i,j\leqslant n$.

我们约定，后面涉及的时态元素，其表达是规范的.

显然，$\forall\ I\in P$，I 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 $u,v,u_1,v_1\cdots$ 来表示时态元素. 时态元素 u 和 v 的并和交分别为 $u+v$ 和 $u\star v$，且 u 的补 $(C-u)$ 表示为 $-u$. 它们的结果也是时态元素.

于是，我们有如下性质.

性质 1. 设 C 上所有的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 $\star$，$+$，$-$ 运算下是一个布尔代数，或是封闭的.

定义 6. 设 u 和 v 是时态元素.

① u 和 v 的差为 $u-v=\{t\,|\,t\in u\wedge t\notin v\}$；

② u 中的最大和最小时刻分别记为 $last(u)$ 和 $first(u)$；

③ $fs(u)$和 $ls(u)$分别是 u 中包含 $first(u)$和 $last(u)$的最大区间.

例如，设 $u=[0,5]+[8,8]$，$v=[3,10]$，则 $first(u)=0$，$last(u)=8$，$first(v)=3$，$last(v)=10$，$fs(u)=[0,5]$，$ls(u)=[8,8]$，$fs(v)=ls(v)=[3,10]$

现在，我们引进模糊时态赋值的观念. 假设一个模糊时态关系的属性 A 的值域 $dom(A)$ 的元素是不包模糊词的，这可以简洁描述.

定义 7. 设 A 是一个属性，它的值域记为 $dom(A)$. A 的一个模糊时态赋值 ζ 是从 u 到 $dom(A)\times[0,1]$的一个函数，它使得每一个 $t\in u$，$\zeta_{(t)}=(a,b)$，其中 $a\in dom(A)$，$b\in[0,1]$，这里，$u\in C$ 是时态域.

与上一节讨论的一样，一个时态赋值中数据的非确定值是一个与时间有关的函数，并假定是阶梯函数.

由于我们的关系是 $N1NF$ 的，所以上面定义中的模糊时态赋值 ζ 是一个集合. 为了方便，我们将 ζ 表示为 $\zeta=\{\eta_1,\eta_2,\cdots,\eta_m\}=\{\langle v_1,a_1,b_1\rangle,\langle v_2,a_2,b_2\rangle,\cdots,\langle v_m,a_m,b_m\rangle\}$这里，$\eta_i$ 是 ζ 的元素，v_i 是时态元素，$a_i\in dom(A)$，$b_i\in[0,1]$，$1\leqslant i\leqslant m$，且 $\zeta_{(t)}=(a_i,b_i)$，如果 $t\in v_i$，$1\leqslant i\leqslant j$；以及如果 $i\neq j$，则 $a_i=a_j$ 和 $b_i=b_j$ 不同时成立，$1\leqslant i,j\leqslant m$.

注意，上述的 $m\geqslant 1$.

进一步，我们用$\overline{\zeta}$表示 ζ 的时态域，即$\overline{\zeta}$是 $v_1,v_2,\cdots,v_m$ 的并；$|\zeta|$表示 ζ 的范围，$|\zeta|=\{(a_1,b_1),(a_2,b_2),\cdots,(a_m,b_m)\}$. 对于任意的 $\eta_i\in\zeta$，我们用$[\eta_i]$表示 a_i，用$\overline{\eta_i}$表示 b_i，$1\leqslant i\leqslant m$. 如果 v 是一个时态元素，我们用 $\zeta\uparrow v$ 表示 ζ 在 $v*\overline{\zeta}$上的限制. 显然，限制 $\zeta\uparrow v$ 也可以是 A 的一个模糊时态赋值.

定义 8. 设 ζ_1 和 ζ_2 是属性 A 的时态赋值.

① 如果$\forall\, t\in\overline{\zeta_1}*\overline{\zeta_2}(\|\zeta_{1(t)}-\zeta_{2(t)}\|<\varepsilon)$，称 ζ_1 和 ζ_2 一致，则 $\zeta_1\cup\zeta_2$ 是 ζ_1 和 ζ_2 到时态域 $\overline{\zeta_1}+\overline{\zeta_2}$上的一个一般扩充，该扩充也是 A 的时态赋值.

② 设 $u=\{t\mid t\in\overline{\zeta_1}*\overline{\zeta_2}\wedge\|\zeta_{1(t)}-\zeta_{2(t)}\|<\varepsilon\}$. 我们定义 $\zeta_1\cap\zeta_2$ 为 $\zeta_1\uparrow u$(或 $\zeta_2\uparrow u$)，且 $\zeta_1-\zeta_2=\zeta_1\uparrow(\overline{\zeta_1}-u)$.

下面，我们给出模糊时态关系的一些定义和规定.

定义 9. 一个关系模式 R 上的元组 τ 是 R 上的这样一个函数，它使得 R 的每个属性 A，$\tau(A)$是 A 的一个模糊时态赋值. R 上的元组 τ，若$\forall\, A\in R$，$\overline{\tau(A)}=\Phi$，则 τ 是空值.

我们假设后面讨论的模糊时态关系中没有空值元组.

定义 10. 设 R 是一个关系模式，给定 $K\subseteq R$ 是 R 上的关键字集，那么 R 上的一个关系 r 是上述有限个非空元组的集合，它满足：

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

② 对于$\forall\,\tau_1,\tau_2\in r$，$\forall\, A\in K(\||\tau_1(A)|-|\tau_2(A)|\|<\varepsilon)$当且仅当 $\tau_1=\tau_2$，$\varepsilon>0$ 为某个足够小的正数.

通常，定义 10 中 r 应满足的 2 个条件被称为模糊时态范式.

3　关系代数

为简化后面的模糊时态关系代数操作的定义，我们需要增加定义 2 个操作符，它们是同

键值归并“⊕”和元组分取“①”两个运算符.

(1)同键值归并⊕操作:对于关系模式 R 上的2个关系 r' 和 r''. 设 $K\subseteq R$ 为关键字集, $A=R-K$ 为非关键字集合,并且 r' 和 r'' 是一致的,即元组间一致无矛盾. 将 r' 和 r'' 合并为 r 时,如果 $\tau'\in r'$ 和 $\tau''\in r''$ 在 K 上有,对于 $K'\in K(\|\,|\tau'(k')|-|\tau''(k')|\,\|<\varepsilon)$, $\varepsilon>0$ 为某个小数,则 τ' 和 τ'' 合并为 r 中的元素 τ 时需要用⊕操作. 经⊕操作后, τ 为:

① $\forall k'\in k, |\tau(k')|=|\tau'(k')|, \overline{\tau(k')}=\overline{\tau'(k')}+\overline{\tau''(k')}$;

② 对 $\forall A'\in A$, 设 $F=\tau'(A')=\{\langle u_1,a_1,b_1\rangle,\cdots,\langle v_l,c_l,d_l\rangle\}$, $H=\tau''(A'')=\{\langle v_1,c_1,d_1\rangle,\cdots,\langle v_k,c_k,d_k\rangle\}$,则

a)对于任意的 $1\leqslant i\leqslant l$,若存在一个 $J(1\leqslant J\leqslant k)$,使得 $a_i=c_j$ 且 $\|b_i-d_j\|<\varepsilon$,则将 $\langle u_i+v_j,a_i,b_i\rangle$ 增加到 G 中,并将 $\langle u_i,a_i,b_i\rangle$ 和 $\langle v_j,c_j,d_j\rangle$ 分别从 F 和 H 中去掉;

b)对于任意 $1\leqslant x\leqslant l$, $a_x\notin\{c_1,c_2,\cdots,c_K\}$,将 $\langle u_X,a_X,b_X\rangle$ 增加到 G 中并删除 F 中该元素;

c)对于任意 $1\leqslant z\leqslant k$, $c_z\notin\{a_1,a_2,\cdots,a_l\}$,将 $\langle v_z,c_z,d_z\rangle$ 增加到 G 中并删除 H 中该元素;

d)将 F 和 H 中剩下的元素增加到 G 中;

e)若存在 $\langle u',a',b'\rangle$ 和 $\langle u'',a'',b''\rangle\in G$,且 $last(u')=last(u'')=NOW$. 根据上面的操作,显然, $a'\neq a''$. 如果 $first(ls(u'))<first(ls(u''))$,则用 $[first(ls(u')),first(ls(u''))-1]$ 代替 G 中的 $\langle u',a',b'\rangle$ 中 u' 的 $ls(u')$;如果 $first((ls(u'))>first(ls(u''))$,则用 $[first(ls(u'')),first(ls(u'))-1]$ 代替 $\langle u'',a'',b''\rangle$ 中 u'' 的 $ls(u'')$;如果 $first(ls(u'))=first(ls(u''))$,则与 r' 和 r'' 一致相矛盾;

f) $\tau(A')=G$;我们记 τ_1 和 τ_2 的⊕操作为 $\tau_1\oplus\tau_2$.

(2)元组分取①操作:元组分取①操作的要求和假设与操作相同.

下面定义 $\tau=\tau'①\tau''$.

① $u=\{t\,|\,B\in R\wedge t\in\overline{\tau'(B)}\times\overline{\tau''(B)}\wedge(\|\,|K'(B)\uparrow[t,t]|-|\tau''(B)\uparrow[t,t]|\,\|)\}$;

② $\tau=\tau'\uparrow u$.

4　基本的模糊时态关系代数操作

基本的模糊时态关系代数操作是传统的基本关系代数操作的扩展.

1)扩展并($\cup^{ft}$)

设 r_1 和 r_2 分别为模糊时态关系模式 R_1 和 R_2 上的关系. 若进行 r_1 和 r_2 的并运算,要求 $R_1=R_2$,及 r_1 和 r_2 一致,且有相同的关键字集. 设 r_1 和 r_2 的并为 r, r 是 $R=R_1=R_2$ 上的一个关系,且 $K\subseteq R$ 是 R 的关键字集, $A=R-K$ 是 R 的非关键字集. 那么,($\varepsilon>0$ 为某个小数)

$$r=r_1\cup^{ft}r_2=\{\tau\,|\,(\exists\,\tau_1\in r_1,\exists\,\tau_2\in r_2(K'\in k(\|\,|\tau_1(K')|-|\tau_2(K')|\,\|<\varepsilon)\wedge\tau=(\tau_1\oplus\tau_2))\vee(\tau\in r_1\wedge(\forall\,\tau'\in r_2(\exists\,K'\in K(\|\,|\tau(K')|-|\tau'(K')|\,\|>\varepsilon)))\vee(\tau\in r_2\wedge(\forall\,\tau'\in r_1(\exists\,K'\in K(\|\,|\tau(K')|-|\tau'(K')|\,\|>\varepsilon))))\}$$

2)扩展差($-^{ft}$)

扩展差的要求与扩展并的要求(或假设)相同. 那么

$$r=r_1-^{ft}r_2=\{\tau\,|\,(\exists\,\tau_1\in r_1,\exists\,\tau_2\in r_2(\forall\,K'\in k(\|\,|\tau_1(K')|-|\tau_2(K')|\,\|<\varepsilon)\wedge(\tau_{1(R)}①\tau_{2(R)}\neq\Phi\wedge\tau=(\tau_1\uparrow(\overline{\tau_1}-\overline{\tau_1①\tau_2}))\uparrow(\overline{\tau_1}\theta ls(\overline{\tau_2})))\vee(\tau_{1(R)}①\tau_{2(R)}=\Phi\wedge\tau=\tau_1\uparrow(\overline{\tau_1}\theta ls(\overline{\tau_2})))))$$

$\vee(\tau \in r_1 \wedge (\forall \tau' \in r_2(\exists K' \in K(\| |\tau(K')| - |\tau'(K')| \| > \varepsilon))))\}$.

注意，上式中的“$-$”号的意义与第 2 节定义的时态元素差一致. 而“θ”需要在下面进行补充定义.

定义 11. 设 u_1 和 u_2 为 2 个时态元素，我们定义“θ”操作如下：

① $NOW \in ls(u_1)$，$NOW \in ls(u_2)$，且 $ls(u_1) \supset ls(u_2)$，则 $u_1 \theta u_2 = (u_1 - ls(u_1)) + (ls(u_1) - ls(u_2))$；

② 其它情况下，$u_1 \theta u_2 = u_1$.

操作符 θ 主要用于扩展差，使之与扩展并一样同时考虑了 NOW 的处理. 当然，对 NOW 的简单处理是不够的，需要单独对 NOW 作研究. 因篇幅有限，在此省略.

3)扩展交($\cap^{ft}$)

扩展交的要求和假设与扩展并相同. 那么

$r = r_1 \cap^{ft} r_2 = \{\tau | (\exists \tau_1 \in r_1, \exists \tau_2 \in r_2(\forall K' \in k(\overline{\tau_1(K')} \times \overline{\tau_2(K')} \neq \Phi \wedge \| |\tau_{1(K')} - |\tau_{2(K')}| \|) < \varepsilon \wedge (\forall A' \in A(\tau_{1(A')} ① \tau_{2(A')} \neq \Phi)) \wedge \tau = \tau_1 ① \tau_2)))\}$.

4)扩展笛卡尔积(X^{ft})

设 r_1 和 r_2 分别为模糊时态关系模式 R 和 S 上的关系. 与传统笛卡尔积的假设一样，r 是 RS 上的一个关系，并且 r 的元组 $\tau = (a_1, \cdots, a_n, b_{n+1}, \cdots, b_{n+m})$，其中$(a_1, \cdots, a_n)$是 r_1 的元组，$(b_{n+1}, \cdots, b_{n+m})$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 $K_1 \subseteq R$，$K_2 \subseteq S$ 分别是 R 和 S 的关键字集，则 K_1K_2 是 RS 的关键字集. 那么，

$r = r_1 X^{ft} r_2 = \{\tau | \tau = (\tau_1, \tau_2), \tau_1 \in r_1, \tau_2 \in r_2\}$.

5)扩展自然连接($\Join^{ft}$)

设 r_1 和 r_2 分别为模糊时态关系模式 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 $R = (A, X, B)$就是连接结果关系 r 的模糊时态关系模式，r 应满足：

$r = r_1 \Join^{ft} r_2 = \{\tau | (\forall \tau_1 \in r_1, \wedge \exists \tau_2 \in r_2(\overline{\tau_1} \times \overline{\tau_2} \neq \Phi \wedge u = \{t \| \tau_{1(t)} - t_{2(t)} \| < \varepsilon\} \wedge u \neq \Phi \wedge \tau_{(A)} = \tau_{1(A)} \uparrow u \wedge \tau_{(B)} = \tau_{2(B)} \uparrow u \wedge \tau_{(X)} = \tau_{1(X)} \uparrow u))\}$.

设 K_1 和 K_2 分别是 R_1 和 R_2 的关键字集，则 R 的关键字集为 $K = K_1 \cup K_2$.

6)扩展投影(Π^{ft})

扩展投影是使用扩展并 $\cup^{ft}$ 操作将传统投影结果合并成一个关系. 设 r_1 是模糊时态关系模式 R_1 上的关系，$y \subseteq R_1$，则在 y 上的扩展投影记为 $r = \Pi_y^{ft}(r_1)$，r 是 $R = y$ 上的关系. 那么 $r = \Pi_y^{ft}(r_1) = \cup^{ft}(\Pi_y(\tau))$.

7)扩展选择(σ^{ft})

设 r_1 是模糊时态关系模式 R_1 上的一个关系. 传统的选择操作(σ)是 $r_2 = \sigma_p(r_1)$，其中 p 是谓词，它表达一个限制. 设 Q 是一个时态表达式，则扩展选择为：

$r = \sigma_{p \wedge Q}^{ft}(r_1) = \{\tau | \tau = (\tau_1 \uparrow p) \uparrow Q \vee \tau = \sigma_p(\tau_1) \uparrow Q\}$

这里，Q 需要进行定义. 我们采用递归形式定义时态表达式如下：①每个时态元素 u 是一个时态表达式；②如果 u 和 v 是时态表达式，则 $u+v$，$u*v$，$u-v$，$first(u)$，$last(u)$，$fs(u)$ 和 $ls(u)$是时态表达式；③如果 F 是一个模糊时态元组演算表达式，则$\overline{F}$是一个时态表达式；④模糊时态元组演算表达式为$\{\tau \uparrow \psi(\tau) \hat{} \varphi(\tau)\}$这里，$\varphi$是时间限制，$\psi$ 是一个公式，实际上，

模糊时态元组演算表达式是传统元组演算表达式加上时间限制.

8)模糊时态关系代数操作的封闭性

上面,我们已给出了时态关系代数操作的定义.从定义看,显然有如下定理.

定理 1. 在扩展并、交、差、笛卡尔积、自然连接、投影和选择操作下,模糊时态范式是封闭的.

5 总 结

从上面的讨论可以看出,时态这个因素在数据库中难以正确模拟.因此,从现有的有关文献看,时态数据库的理论离实用还有一段距离.但随着实际要求的增强,这些问题必须解决,这需要大量的热衷于该问题的研究人员努力解决.特别是模糊时态数据库技术刚刚起步,应用前景很广阔.笔者希望借本文抛砖引玉.

致谢 本文大部分工作在数学所访问时完成,感谢陆汝钤老师的精心指导与支持!

参考文献

1 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**):936—939.

2 Mckenzie L, Smodgrass R. Evaluation of relational algebras in corporating the time dimension in databases. ACM Computing Surveys, 1991, **23**(**4**).

3 张师超.时态数据库述评.计算机科学,1992(3).

A FUZZY TEMPORAL RELATIONAL ALGEBRA

Zhang Shichao

(*National Research Center for Intelligent Computing Systems, Beijing* 100080)

(*Department of Mathematics, Guangxi Normal University, Guilin* 541001)

Luo Xudong

(*Department of Automation Engineering, Beijing Institute of Light Industry, Beijing* 100037)

Abstract The ubiquitous nature of time requires that a database management system provide facilities for modelling of time aspect of data in the real world. This paper establishes a fuzzy temporal relational algebra. This algebra avoids a weakness in the current temporal database techniques, which will be pointed out in this paper.

Key words Temporal database, historical database, relational algebra, fuzzy database.