全序时态模块模式的 TO_TSNF 分解问题研究*)

万 静[1] 郝忠孝[1,2,3]

(哈尔滨理工大学计算机科学与技术学院 哈尔滨 150080)[1] (齐齐哈尔大学计算机系 齐齐哈尔 161006)[2]

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)[3]

摘 要 在许多时态数据库应用中，都存在着涉及多时间粒度的约束。但是，具有多时间粒度的时态数据库的设计相当复杂，甚至难以实现，而现实世界中的许多应用所涉及到的时态类型集都能满足全序关系。同时，具有全序时态类型集的全序时态模块模式有着良好的特性，因此本文提出了全序时态模块模式、时刻关系模式、全序时态模块投影、全序时态简单候选关键字、全序时态三范式(TO_T3NF)和全序时态简单范式(TO_TSNF)等概念，并给出了全序时态简单范式的分解算法，对其正确性进行了证明，并对时间复杂度进行了分析。

关键词 全序时态模块模式，全序时态模块投影，全序时态简单范式，多时间粒度

Research on TO_TSNF Decomposition with Total Order Temporal Module Scheme

WAN Jing[1] HAO Zhong-Xiao[1,2,3]

(College of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080)[1]

(Department of Computer, Qiqihar University, Qiqihar 161006)[2]

(College of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)[3]

Abstract Constraints involving multiple time granularities exist in many temporal database applications. However, design of temporal database with multiple time granularities is fairly complicated, even hard to implement. Temporal type sets involved in many real world applications can constitute one total order of types, and total order temporal module schemes with total order temporal type sets have fine properties. The concepts of total order temporal module scheme, tickwise relation scheme, total order temporal module projection, total order temporal simple candidate key, total order temporal 3NF and total order temporal SNF etc. are given in this paper. The total order temporal SNF decomposition algorithm, the proof for its correction and the time complexity analysis are also given in this paper.

Keywords Total order temporal module schemes, Total order temporal module projection, Total order temporal SNF, Multiple time granularities

1 引言

在时态数据库设计中，时态范式起着重要的作用。已经提出的一些时态数据库范式有：Segev 的 1TNF[1]，S. B. Navathe 的 TNF[2]，N. A. Lorentzos 的 P 和 Q 范式[3]，Christian S. Jensen 的 T3NF 与 TBCNF[3]，SEAN Wang 的 T3NF 与 TBCNF[4]。其中，Segev 的 1TNF，S. B. Navathe 的 TNF，N. A. Lorentzos 的 P 和 Q 范式都是针对特定的时态数据模型，而且它们都偏离了传统的关系数据库范式，不是传统范式真正意义上的扩展；Christian S. Jensen 的 T3NF 与 TBCNF 是对传统范式的扩展，但只考虑了时态关系快照中的数据冗余，没有考虑快照间的数据冗余，而且没有考虑多时间粒度的问题；X. S. Wang 的 T3NF 与 TBCNF 是对 Christian S. Jensen 的工作的扩展，考虑了快照间的数据冗余，而且考虑了多时间粒度的问题。但是，由于多时间粒度的引入，使得 X. S. Wang 的 T3NF 与 TBCNF 的分解算法相当复杂，难以用其进行有效的时态数据库设计。而现实世界中的许多应用所涉及到的时态类型集都能满足全序关系，同时具有全序时态类型集的全序时态模块模式有着良好的特性，因此本文基于全序时态模块模式提出了全序时态三范式(TO_T3NF)的概念。由于 TO_T3NF 完全不能消除时态主属性对时态关键字的传递依赖，本文进一步提出了规范程度更高的全序时态简单范式(TO_TSNF)的概念，并给出了全序时态简单范式的分解算法，对其正确性进行了证明，对算法进行了分析。

2 基本概念

对于时态类型、细于关系、时态模块模式与时态模块、时态关键字、时态函数依赖(TFD)、TFD 导出和逻辑蕴涵等概念的定义以及 TFD 推导公理、TFD 有限推导公理的描述均参见文[4]。

定义 1[5](全序时态类型集) 给定时态类型集 $T=\{\mu_1, \mu_2, \cdots, \mu_n\}$，若对于其中任意两个 μ_i，μ_j，$1\leqslant i\leqslant j\leqslant n$，都有 $\mu_i \leqslant \mu_j$ 或 $\mu_j \leqslant \mu_i$，则称 T 是全序时态类型集。

定义 2(全序 TFD 集) 设 F 为一 TFD 集，如果对于 F 中的任意两个 TFD $X \rightarrow_{\mu_1} Y$，$V \rightarrow_{\mu_2} W$，有 $\mu_1 \leqslant \mu_2$ 或 $\mu_2 \leqslant \mu_1$，则称 F 为全序 TFD 集。

定义 3(平凡的 TFD) 对于(R, μ)上成立的 TFD $X \rightarrow_\nu Y$，如果 $Y \subseteq X$，或者 $\nu \leqslant \mu$，那么称 $X \rightarrow_\nu Y$ 是(R, μ)上一个

*)黑龙江省自然科学基金资助项目(F00-06)。**万 静** 博士生，副教授，主要研究方向为时态数据库系统与理论；**郝忠孝** 博士生导师，教授，主要研究方向为数据库系统与理论。

平凡的 TFD。

定义 4(全序时态模块模式) 一个时态模块模式(R, μ)是全序时态模块模式,如果(R, μ)上成立的 TFD 集 F 为全序 TFD 集,且对于任一 TFD $X \rightarrow_{\nu} Y \in F$,有 $\mu \leqslant \nu$。一个全序时态模块是一个三元组(R, μ, ϕ),(R, μ)为全序时态模块模式,ϕ 是一个时间窗口函数,即从 N(正整数)到 $2^{Tup(R)}$ 的映射($Tup(R)$为全部元组的集合),使得对每个 $i \geqslant 1$,如果 $\mu(i)=\phi$,则 $\phi(i)=\phi$。

注:对于(R, μ)上的任一平凡的 TFD $X \rightarrow_{\nu} Y \in F, \nu \leqslant \mu$,其在$(R, \mu)$上将自动成立,且在范式的分解问题中不起任何作用,故不在本文的考虑范围之内。

定义 5(时刻关系模式) 对于一个时态模块模式(R, μ),当 μ 取值某一确定的非空时刻 $\mu(i)$时,$(R, \mu(i))$就构成了一个时刻关系模式。

由于 $\mu(i)$为一个确定的值,故(R, μ)在非空时刻 $\mu(i)$的时刻关系模式$(R, \mu(i))$相当于传统的关系模式 R。

定义 6(全序 TFD 集的模式投影) 给定时态模块模式(R, μ)和全序 TFD 集 F,则 F 到模式(R, μ)上的投影,记作 $\pi_{(R,\mu)}(F)$,定义为:

$\pi_{(R,\mu)}(F)=\{X \rightarrow_{\nu} Y \mid F \mid = X \rightarrow_{\nu} Y, XY \subseteq R$ 且 $\mu \leqslant \nu\}$。

定义 7(全序 TFD 集的非时态版) 设 F 为全序时态模块模式(R, μ)上成立的全序 TFD 集,它的非时态版为:$\pi_{\phi}(F)=\{X \rightarrow Y \mid \exists X \rightarrow_{\nu} Y \in F, \mu \leqslant \nu\}$,即 $\pi_{\phi}(F)$为 F 中所有 TFD 的非时态版本。

定义 8(全序时态模块投影) 全序时态模块 $M=(R, \mu, \phi)$, $R_1 \subseteq R, \mu \leqslant \nu$,则 M 在(R_1, ν)上的投影 $\pi_{(R1,\nu)}(M)=(R_1, \nu, \phi_1)$,对任意 $i \geqslant 0$, $\phi_1(i)=\bigcup_{j:\mu(j) \subseteq \nu(i)} \pi_{R_1}(\phi(j))$,其中 π 为传统关系数据库中的投影操作,ϕ 及 ϕ_1 分别为 M 及 $\pi_{(R1,\nu)}(M)$的时间窗口函数。

定义 9(全序时态三范式,TO_T3NF)) 一个全序时态模块模式(R, μ)是属于全序时态三范式(TO_T3NF)的,如果 F 中的每个 TFD $X \rightarrow_{\nu A}, XA \subseteq R, A \notin X, \mu \leqslant \nu$,以下条件之一成立:

(1) A 是(R, μ)的时态主属性,或者

(2) X 是(R, μ)的时态超候选关键字,且 $\mu=\nu$。

注:如果 F 中的一个 TFD $X \rightarrow_{\nu} A, XA \subseteq R, A \notin X, \mu \leqslant \nu$,那么$(R, \mu)$中一定存在时态冗余,故定义中要求 $\mu=\nu$。

3 全序时态简单候选关键字集算法

定义 10(全序时态简单候选关键字) 设(R, μ)为一全序时态模块模式,F 是(R, μ)上成立的全序 TFD 集。$X \subseteq R$, X 为(R, μ)的时态候选关键字。如果对任意属性 $A \in R(A \notin X)$,不存在 X 的真子集 $X' \subset X$,使得 $X' \rightarrow_{\mu} A$ 被 F 逻辑蕴涵,则称 X 为(R, μ)的全序时态简单候选关键字。全序时态简单候选关键字包含的属性称为全序时态简单主属性。

定理 1 若(R, μ)是一个全序时态模块模式,F 是(R, μ)上成立的全序 TFD 集,那么 $F \mid = X \rightarrow_{\mu} Y$,当且仅当 $\pi_{\phi}(F) \mid = X \rightarrow Y$。

证明:(充分性)设 $\pi_{\phi}(F) \mid = X \rightarrow Y$,则 $X \rightarrow Y$ 能由 $\pi_{\phi}(F)$用 Armstrong[8] 公理导出,仅需证明 $X \rightarrow_{\mu} Y$ 能由 F 用 TFD1～TFD4[4] 导出即可。下面对推导步数作归纳证明。

当 $n=1$ 时,有 $Y \subseteq X$ 或者 $X \rightarrow Y \in \pi_{\phi}(F)$。如果 $Y \subseteq X$,由 TFD1 有 $F \mid = X \rightarrow_{\mu} Y$;如果 $X \rightarrow Y \in \pi_{\phi}(F)$,则由定义 7 知,必存在 $X \rightarrow_{\nu} Y \in F, \mu \leqslant \nu$,由 TFD4 有 $F \mid = X \rightarrow_{\mu} Y$。

设当 $n=k$ 时,归纳结论成立,即任意的 FD $Z \rightarrow W$,能由 $\pi_{\phi}(F)$用 Armstrong 公理用 k 步导出时,有 $F \mid = Z \rightarrow_{\mu} W$。而 $X \rightarrow Y$ 的证明至少需要 $k+1$ 步才能完成。因为 $k \geqslant 1$,只可能有两种情况:一是在前 k 步中由 $\pi_{\phi}(F)$根据 Armstrong 公理已证明了存在某个 $V \subseteq X, V \rightarrow Y$,最后一步应用 FD2 而导出 $X \rightarrow Y$,根据归纳假设,有 $F \mid = V \rightarrow_{\mu} Y$,再根据 TFD2 导出 $F \mid = X \rightarrow_{\mu} Y$;另一种可能是在前 k 步中由 $\pi_{\phi}(F)$根据 Armstrong 公理证明了存在某个 $V, X \rightarrow V$ 且 $V \rightarrow Y$,最后一步应用 FD3 而导出 $X \rightarrow Y$,根据归纳假设有 $F \mid = X \rightarrow_{\mu} V, F \mid = V \rightarrow_{\mu} Y$,再根据 TFD3 可导出 $F \mid = X \rightarrow_{\mu} Y$。因而当 $n=k+1$ 时,在所有两种可能情况下,归纳结论是完全正确的,从而充分性得证。

(必要性)如果能证明若 $F \mid = X \rightarrow_{\nu} Y(\mu \leqslant \nu)$,则 $\pi_{\phi}(F) \mid = X \rightarrow Y$,那么必要性得证。

设 $F \mid = X \rightarrow_{\nu} Y(\mu \leqslant \nu)$,即 $X \rightarrow_{\nu} Y$ 能用 TFD1～TFD4 导出,需证明 $X \rightarrow Y$ 能由 $\pi_{\phi}(F)$用 Armstrong 公理导出。下面对推导步数作归纳证明。

当 $n=1$ 时,有 $Y \subseteq X$ 或者 $X \rightarrow_{\nu} Y \in F(\mu \leqslant \nu)$。如果 $Y \subseteq X$,根据 FD1 有 $\pi_{\phi}(F) \mid = X \rightarrow Y$;如果 $X \rightarrow_{\nu} Y \in F(\mu \leqslant \nu)$,由定义 7 知,$X \rightarrow Y \in \pi_{\phi}(F)$,故有 $\pi_{\phi}(F) \mid = X \rightarrow Y$。

设当 $n=k$ 时,归纳结论成立,即任意的 TFD $Z \rightarrow_{\nu} W(\mu \leqslant \nu)$,能由 F 用 TFD1～TFD4 用 k 步导出时,有 $\pi_{\phi}(F) \mid = Z \rightarrow W$。而 $X \rightarrow_{\nu} Y$ 的证明至少需要 $k+1$ 步完成。因为 $k \geqslant 1$,分为三种情况:①在前 k 步中由 F 根据 TFD1～TFD4 证明存在某个 $V \subseteq X, V \rightarrow_{\nu} Y$,最后一步应用 TFD2 而导出 $X \rightarrow_{\nu} Y$,根据归纳假设,有 $\pi_{\phi}(F) \mid = V \rightarrow Y$,再根据 FD2 导出 $\pi_{\phi}(F) \mid = X \rightarrow Y$;②在前 k 步中由 F 根据 TFD1～TFD4 证明存在某个 V, $X \rightarrow_{\nu} V$ 且 $V \rightarrow_{\nu} Y$,最后一步应用 TFD3 而导出 $X \rightarrow_{\nu} Y$,根据归纳假设有 $\pi_{\phi}(F) \mid = X \rightarrow V, \pi_{\phi}(F) \mid = V \rightarrow Y$,再根据 FD3 可导出 $\pi_{\phi}(F) \mid = X \rightarrow Y$;③在前 k 步中由 F 根据 TFD1～TFD4 导出 $X \rightarrow_{\nu 1} Y, \nu \leqslant \nu 1$,最后一步应用 TFD4 导出 $X \rightarrow_{\nu} Y$,根据归纳假设,在第 k 步直接可得出 $\pi_{\phi}(F) \mid = X \rightarrow Y$。因而当 $n=k+1$ 时,在所有三种可能情况下,归纳结论是完全正确的,从而必要性得证。证毕。

推论 1 若(R, μ)是一个全序时态模块模式,F 是(R, μ)上成立的全序 TFD 集,那么 X 是(R, μ)关于 F 的时态候选关键字,当且仅当 X 是时刻关系模式$(R, \mu(i))$关于$(\phi(F)$的候选关键字。

定理 2[8] 设 $R\langle U, F\rangle$为一关系模式,U 为属性集,F 为函数依赖集。K 为 $R\langle U, F\rangle$的一个非空的候选关键字集。$R\langle U, F\rangle$含有不属于 K 的候选关键字的充要条件是 K 中有一个元素 k_1,F 中有一个 FD: $L \rightarrow A$,使 $L \cup (k_1 - A)$不包含 K 中的任何元素。

算法 1 TS_Candidate-keys(全序时态简单候选关键字集算法)

输入:全序时态模块模式(R, μ),(R, μ)上成立的全序 TFD 集 F。

输出:(R, μ)的全序时态简单候选关键字集 TS_K。

```
TS_Candidate—keys(R, μ, F)
begin
(1) G:=π_φ(F);{求全序时态候选关键字集}
TS_K:=KEYFINDING(R, G);
    for 每一个 K∈TS_K do
    for 每一个 FD: L→A∈G do
        [ test:=true;
```

$S := L \cup (K-A)$;

for 每一个 $J \in$ TS_K and test do

if $J \subseteq S$ then test := false;

if test then TS_K: = TS_K ∪ {消去 S 中的冗余属性};]

(2) {求全序时态简单候选关键字集}

for 每一个 $K \in$ TS_K do

[test: = true;

for 每一个 $A \in R-K$ and test do

for 每一个 $B \in K$ and test do

if MEMBERSHIP(G, $K-B \to A$) then test: = false;

if not test then TS_K: = TS_K − K;]

(3) 输出 TS_K

end.

定理 3 算法 TS_Candidate-keys 是可终止的,并能正确求出全序时态模块模式(R, μ)的全序时态简单候选关键字集 TS_K,其时间复杂度为 $O(n^2p+pq+q^2+npq^2)$。

证明:(终止性)由于 R 中的属性个数、候选关键字的个数以及 F 中 TFD 的个数都是有限的,故算法中的 for 循环都可终止。

(正确性)由定理 2,步骤(1)能正确求出 R 关于 $\pi_\phi(F)$ 的候选关键字集,再根据推论 1 可证明此候选关键字集即为(R,μ)关于 F 的时态候选关键字集。步骤(2)是根据全序时态简单候选关键字的定义,从步骤(1)所求出的全序时态候选关键字集中剔除每一个不满足全序时态简单候选关键字定义的关键字,因此最终所求结果即为全序时态简单候选关键字集,其中 KEYFINDING 和 MEMBERSHIP 算法为传统关系数据库中的求关键字和成员籍算法[8]。

时间复杂度:步骤(1)中,由于 G 中所含的 FD 的个数不会多于 F 中 TFD 的个数,因此 KEYFINDING 算法的时间复杂度为 $O(pq+q^2)$,其中,p 为 F 中 TFD 的个数,q 为 R 中的属性个数,而 for 循环需 $O(n^2p)$ 级,其中,n 为全部候选关键字的个数。故步骤(1)的时间复杂度为 $O(n^2p+pq+q^2)$。

步骤(2)中,MEMBERSHIP 算法的时间复杂度为 $O(p)$,因此步骤(2)的时间复杂度为 $O(npq^2)$。故算法总的时间复杂度为 $O(n^2p+pq+q^2+npq^2)$。

4 全序时态简单范式

一个全序时态模块模式分解成属于 TO_T3NF 的子模式时,完全不能消除时态主属性对时态关键字的传递依赖。因此本文提出了规范程度高于 TO_T3NF 的一个新的范式,并给出了它的一个满足保持函数依赖与无损连接性的模式分解算法。

定义 11(全序时态简单范式,TO_TSNF)) 具有全序 TFD 集 F 约束的全序时态模块模式(R, μ)是属于全序时态简单范式(*TO_TSNF*)的,如果 F 中的每个 *TFD* $X \to_\nu A$, $XA \subseteq R$, $A \notin X$, $\mu \leqslant \nu$,以下条件之一成立:

(1) A 是(R, μ)的全序时态简单主属性,或者

(2) X 是(R, μ)的全序时态超简单候选关键字,且 $\mu=\nu$。

定义 12(全序时态自然联接) 设 $M_1=(R_1, \mu, \phi_1)$, $M_2=(R_2, \nu, \phi_2)$,其中 $\mu \leqslant \nu$。M_1 和 M_2 的全序时态自然联接 $M_1 \bowtie_{TO\_T} M_2$,是时态模块 $M=(R_1 \cup R_2, \mu, \phi)$,$\phi$ 定义如下:对每个 $i \geqslant 1$, $\phi(i)=\phi_1(i) \bowtie \phi_2(j)$, $j \geqslant 1$, $\mu(i) \subseteq \nu(j)$,其中 $\bowtie$ 为传统的自然联接操作。

定义 13(全序时态模块模式分解) 全序时态模块模式(R, μ)的分解是一个时态模块模式集合 $\rho=\{(R_1, \mu_1), \cdots, (R_k, \mu_k)\}$,满足:

(a) $R=R_1 \cup \cdots \cup R_k$

(b) $\{\mu, \mu_1, \cdots, \mu_k\}$ 为一全序时态类型集,且 $\mu \leqslant \mu_i$, $i=1, \cdots, k$。

定义 14(全序无损分解) 设(R, μ)为一全序时态模块模式,F 为全序 TFDs 集。称(R, μ)关于 F 的一个分解 $\rho=\{(R_1, \mu_1), \cdots, (R_k, \mu_k)\}$为全序无损分解,如果对$(R, \mu)$上的每个满足 F 中全部 TFDs 的时态模块 M,有:$M=\pi_{(R1,\mu1)}(M) \bowtie_{TO\_T} \cdots \bowtie_{TO\_T} \pi_{(R_k,\mu_k)}(M)$。

定义 15(全序时刻无损分解) 设(R, μ)为一全序时态模块模式,F 为全序 TFDs 集。称(R, μ)关于 F 的一个分解 $\rho=\{(R_1, \mu_1), \cdots, (R_k, \mu_k)\}$为全序时刻无损分解,如果对 μ 的每个非空时刻 $T \geqslant 1$,下式成立:$\phi(T)=\phi_1(T_1) \bowtie \cdots \bowtie \phi_k(T_k)$,其中 $T_i \geqslant 1$, $1 \leqslant i \leqslant k$,使 $\mu(T) \subseteq \mu_i(T_i)$。

定义 16(保持依赖) 设(R, μ)为一全序时态模块模式,F 为(R, μ)上成立的全序 TFDs 集。称分解 $\rho=\{(R_1, \mu_1), \cdots, (R_k, \mu_k)\}$保持 F 中的依赖,如果 $\bigcup_{i=1}^{k} \pi_{(Ri,\mu i)}(F) \mid = F$。

定理 4 如果一个全序时态模块模式$(R, \mu) \in$ TO_TSNF,则必有$(R, \mu) \in$ TO_T3NF;反之不成立。

证明:根据全序时态简单关键字的定义可知,一个全序时态简单关键字必为时态关键字,而一个时态关键字不一定是全序时态简单关键字;同理,一个全序时态简单主属性必为时态主属性,而一个时态主属性不一定是全序时态简单主属性。因此,满足 TO_TSNF 定义的范式必定满足 TO_T3NF 的定义,反之则不成立。证毕。

TO_T3NF 完全不能消除时态主属性对时态关键字的传递依赖,而 TO_TSNF 则能消除一部分时态主属性对时态关键字的传递依赖,因此其规范化程度要比 TO_T3NF 高。

定理 5 设(R, μ)为一全序时态模块模式,F 为全序 TFDs 集,ρ 为(R, μ)关于 F 的一个分解,则 ρ 为(R, μ)关于 F 的一个全序无损分解,当且仅当它是(R, μ)关于 F 的一个全序时刻无损分解。

证明:(充分性)设 $M=(R, \mu, \varphi)$为一满足 F 的全序时态模块,$\rho=\{(R_1, \mu_1), \cdots, (R_k, \mu_k)\}$为$(R, \mu)$关于 F 的一个全序时刻无损分解。对每个 $1 \leqslant i \leqslant k$,设 $M_i=\pi_{(Ri,\mu_i)}(M)=(R_i, \mu_i, \phi_i)$, $\bigcup_{i=1}^{k} R_i=R$。设 $M'=M_1 \bowtie_{TO\_T} \cdots \bowtie_{TO\_T} M_k=(\bigcup_{i}^{k} R_i, \nu, \phi')=(R, \nu, \phi')$,因为 ρ 为(R, μ)关于 F 的一个全序时刻无损分解,因此对于 μ 的每个非空时刻 $T \geqslant 1$, $\phi(T)=\phi_1(T_1) \bowtie \cdots \bowtie \phi_k(T_k)$,其中 $T_i \geqslant 1$, $1 \leqslant i \leqslant k$,使 $\mu(T) \subseteq \mu_i(T_i)$。由定义 12 可知,$\phi$ 即为 ϕ' 而 $\nu=\mu$,因此 $M=M'=M_1 \bowtie_{TO\_T} \cdots \bowtie_{TO\_T} M_k$,即 ρ 为(R, μ)关于 F 的一个全序无损分解。充分性得证。

(必要性)设 $\rho=\{(R_1, \mu_1), \cdots, (R_k, \mu_k)\}$为$(R,\mu)$关于 F 的一个全序无损分解,$M=(R, \mu, \phi)$为一个满足 F 的全序时态模块。对每个 $1 \leqslant i \leqslant k$,设 $M_i=\pi_{(Ri,\mu_i)}(M)=(R_i, \mu_i, \phi_i)$, $\bigcup_{i=1}^{k} R_i=R$。由 ρ 为(R, μ)关于 F 的一个全序无损分解可知,$M=M_1 \bowtie_{TO\_T} \cdots \bowtie_{TO\_T} M_k$,再根据定义 12 可知,对于 μ 的每个非空时刻 T, $\phi(T)=\varphi_1(T_1) \bowtie \cdots \bowtie \phi_k(T_k)$,其中 $T_i \geqslant 1$, $1 \leqslant i \leqslant k$,使 $\mu(T) \subseteq \mu_i(T_i)$。即 ρ 为(R, μ)关于 F 的一个全

序时刻无损分解。必要性得证。证毕。

定理 6[8] 设 $R\langle U,F\rangle$ 为一关系模式，U 为属性集，F 为函数依赖集。$\rho=\{R_1\langle U_1,F_1\rangle,\cdots,R_n\langle U_n,F_n\rangle\}$ 为 $R\langle U,F\rangle$ 的一个保持 FD 的分解。如果某个 $R_i(1\leqslant i\leqslant n)$ 含有 $R\langle U,F\rangle$ 的候选关键字 K，则 ρ 必是无损连接分解。

定理 7 给定全序时态模块模式 (R,μ)，全序 TFD 集 F 和 (R,μ) 的一个保持依赖的分解 $\rho=\{(R_1,\mu_1),\cdots,(R_k,\mu_k)\}$，如果存在 $(R_i,\mu_i)\in\rho,1\leqslant i\leqslant k$，满足 $\mu_i=\mu$，且 R_i 包含 (R,μ) 的一个时态关键字，则 ρ 一定是一个全序无损的分解。

证明：根据定理 5，只需证明 ρ 是一个全序时刻无损分解，即可证明 ρ 是一个全序无损分解。对于 μ 的任一非空时刻 T，由定义 13 知，在 μ_j 中都存在且仅存在一个时刻 $T_j(1\leqslant j\leqslant k)$，使得 $\mu(T)\subseteq\mu_j(T_j)$。因此在 μ 的任一非空时刻 T，$\phi(T)$ 和 $\phi_j(T_j)(1\leqslant j\leqslant k)$ 就分别构成了传统的关系实例，其关系模式分别为 R 和 $R_j(1\leqslant j\leqslant k)$。因为存在 $(R_i,\mu_i)\in\rho,1\leqslant i\leqslant k$，满足 $\mu_i=\mu$，且 R_i 包含 (R,μ) 的一个时态关键字，根据推论 1，此时态关键字也是时刻关系模式 (R,T) 关于 $\pi_\phi(F)$ 的关键字，即可以理解为此时态关键字是 T 时刻的 R 的关键字。因为 $\rho=\{(R_1,\mu_1),\cdots,(R_k,\mu_k)\}$ 是 (R,μ) 的一个保持依赖的分解，即 $\bigcup_{i=1}^{k}\pi_{(R_i,\mu_i)}(F)\mid=F$，根据定理 1，$\bigcup_{i=1}^{k}\pi_\phi(\pi_{(R_i,\mu_i)}(F))\mid=\pi_\varphi(F)$，可知 $\rho'=\{R_1,\cdots,R_k\}$ 是 T 时刻的 R 的一个保持依赖的分解，且其中的一个 $R_i(1\leqslant i\leqslant k)$ 包含 R 的一个关键字，由定理 6 可知，ρ' 为无损联接分解，即 $\phi(T)=\phi_1(T_1)\bowtie\cdots\bowtie\phi_k(T_k),1\leqslant i\leqslant k$。由此可证 ρ 为一个全序时刻无损分解。证毕。

引理 1 设 (R_i,μ_i) 是 (R,μ) 关于全序 TFD 集 F 的分解 ρ 中的一个模式，其中 $R_i=XA$，$X\rightarrow_{\mu_i}A\in$ TO_CANONICAL(F)[5,6]，$\mu\leqslant\mu_i$，则 X 是 (R_i,μ_i) 的全序时态简单候选关键字。

证明：由于 $X\rightarrow_{\mu_i}A\in$ TO_CANONICAL(F)，因此不存在 $X'\subset X$，使得 $F\mid=X'\rightarrow_{\mu_i}A$，所以 X 是 (R_i,μ_i) 的时态候选关键字。又因为 A 是 R_i 中唯一不属于 X 的属性，根据定义 10 可知 X 是 (R_i,μ_i) 的全序时态简单候选关键字。证毕。

引理 2 设 (R_i,μ_i) 是 (R,μ) 关于全序 TFD 集 F 的分解 ρ 中的一个模式，其中 $R_i=XA$，$X\rightarrow_{\mu_i}A\in$ TO_CANONICAL(F)，$\mu\leqslant\mu_i$，则 (R_i,μ_i) 是全序时态简单范式。

证明：(反证法)。设 $\pi_{(R_i,\mu_i)}(F)\mid=W\rightarrow_\nu A$，$WA\subseteq R_i$，$A\notin W$，$\mu_i\leqslant\nu$，且 W 不是 (R_i,μ_i) 的全序时态简单候选关键字，由引理 1 可知 X 是 (R_i,μ_i) 的全序时态简单候选关键字，则 $W\subset X$。由 $F\mid=W\rightarrow_\nu A$ 和 $F\mid=X\rightarrow_{\mu_i}A$ 可知 $X\rightarrow_{\mu_i}A$ 是冗余的，这与 $X\rightarrow_{\mu_i}A\in$ TO_CANONICAL(F) 相矛盾，所以 W 一定是 (R_i,μ_i) 的全序时态简单候选关键字 X，根据定义 11，(R_i,μ_i) 是全序时态简单范式。证毕。

算法 2 TO_TSNF_DECIDING(TO_TSNF 判定算法)

输入：全序时态模块模式 (R,μ)，(R,μ) 上成立的全序 TFD 集 F。

输出：若 (R,μ) 属于 TO_TSNF，则输出 true；否则输出 false。

```
TO_TSNF_DECIDING(R, μ, F)
begin
 test:=true;
 TS_K:=TS_Candidate—keys(R, μ, F);
 for 每一个 TFD:X→νA∈F and test do
   if μ≠ν then test:=false;
 for 每一个 TFD:X→νA∈F and test do
   [ flag:=false;
     for 每一个 K∈TS_K and not flag do
       if K⊆X then flag:=true;
     if not flag then for 每一个 K∈TS_K do
       if A∈K then flag:=true;
     if not flag then test:=false;]
   return(test);
 end.
```

定理 8 算法 TO_TSNF_DECIDING 是可终止的，并能正确判定全序时态模块模式 (R,μ) 是否属于全序时态简单范式，其时间复杂度为 $O(n^2p+pq+q^2+npq^2+np+p)$。其中 n，p 和 q 的含义同算法 1。

证明：(终止性)由于算法 TS_Candidate-keys 是可终止的，并且由于 F 中 TFD 的个数和 (R,μ) 中全序时态简单候选关键字的个数都是有限的，故算法中的 for 循环可终止。

(正确性)算法根据全序时态简单范式的定义，对于 F 中的每一个 TFD，先测试它的时态类型是否与全序时态模块模式 (R,μ) 的时态类型一致，再测试它的左部是否包含全序时态简单候选关键字，如果不包含，再测试它的右部是否为全序时态简单主属性，只有满足两者条件之一的，才判定为全序时态简单范式，故算法是正确的。

时间复杂度：算法 TS_Candidate—keys 的时间复杂度为 $O(n^2p+pq+q^2+npq^2)$，第一个 for 循环需 $O(p)$ 级，第二个 for 循环需 $O(np)$ 级，故算法总的时间复杂度为 $O(n^2p+pq+q^2+npq^2+np+p)$。

算法 3 TO_TSNF(全序时态简单范式分解算法)

输入：全序时态模块模式 (R,μ)，(R,μ) 上成立的全序 TFD 集 F。

输出：满足联接无损性、保持依赖的 TO_TSNF 数据库模式 ρ。

```
TO_TSNF(R, μ, F)
begin
(1)F :=F ∪{R→μC};{其中 C 是任一在 R 中不出现的
                        虚属性名}
  G:=TO_CANONICAL(F);
(2) 将 G 中 TFD 按左部相同和时态类型相同归类;{合
    并满足 TO_TSNF 条件的 TFD}
 for G 中每一个左部 Xi 和时态类型 νj do
   G(Xi, νj):={Ak | G|=Xi→νjAk};
 for 每一组 G(Xi, νj)do
   [ n:=G(Xi, νj)中元素个数; G':=G(Xi, νj);
     while G'≠φ do
     [ change:=false; S:=n−1;
       G':=G'−A1; Y:=A1;
       for K:=2 to n do
         [ H :=π(XiYAk,νj)(G);
           if TO_TSNF_DECIDING
             (Xi∪Ak, νj, H)then
               [ Y :=YAk;
                 G':=G'−Ak;
                 S:=S−1;
                 change:=true;]]
```

if change then 用 $X_i \rightarrow_{\nu j} Y$ 替换 G 中相应 TFD $X_i \rightarrow_{\nu j} A_1 \cdots X_i \rightarrow_{\nu j} A_m$;

$n := S$;

将 G' 中的 A_i 重新从 1 到 n 编号;]]

(3)$\rho := \phi$;

for 每一个 $X \rightarrow_{\nu} Y \in G$ do

$\rho := \rho \cup (XY, \nu)$;

(4)删除第一步引入的虚属性}

for 任意一个$(X, \nu) \in \rho$ do

if $C \in X$ then $\rho := \rho - (X, \nu) \cup (X - C, \nu)$

(5)change := true;

while change do

[change := false;

for 任意两个(X, ν),$(Y, \nu) \in \rho$ do

[if $X \subseteq Y$ then[$\rho := \rho - (X, \nu)$;

change := true;]

else if $Y \subseteq X$ then[$\rho := \rho - (Y, \nu)$;

change := true;]]]

(6) return(ρ);

end.

定理 9 算法 TO_TSNF 正确求得了全序时态模块模式(R, μ)关于全序 TFDs 集 F 的一个满足保持依赖、无损连接的 TO_TSNF 的分解 ρ,其时间复杂度为 $O(p^2q^2 + pq^3m)$,p,q 和 m 分别为 F 中的 TFD 的个数、R 中属性个数以及 F 中时态类型的个数。

证明:(终止性)由于算法 TO-CANONICAL 是正确的[5],并且 G 中包含有限个 TFD,显然算法是可终止的。下面证明正确性。

(1)TO_TSNF 满足性

算法最终所得到的分解 $\rho = \{(R_1, \mu_1), \cdots, (R_k, \mu_k)\}$,其中的模式$(R_i, \mu_i)$ $(1 \leqslant i \leqslant k)$可分为三种类型:

① $R_i = XA$, $X \rightarrow_{\mu i} A \in$ TO_CANONICAL(F);

② $R_i = XA_1 \cdots A_m$, $X \rightarrow_{\mu i} A_i \in$ TO_CANONICAL(F), $1 \leqslant i \leqslant m$;

③ R_i 为(R, μ)的时态关键字。

由引理 2 知①中的(R_i, μ_i)是全序时态简单范式;而③中的 R_i 即为(R_i, μ_i)的全序时态简单候选关键字,因此③中的(R_i, μ_i)也是全序时态简单范式;②中的 R_i 是经过 TO_TSNF 判定算法 TO_TSNF_DECI- DING 判定后才合并相应的 $X \rightarrow_{\mu i} A_i$,因此$(R_i, \mu_i)$必为全序时态简单范式。由此得证 ρ 满足 TO_TSNF。

(2)保持依赖性

由于 $F \equiv G$,而根据算法的第(2)步和第(3)步,G 中的每一个 TFD $X \rightarrow_{\nu} A$,其左部和右部属性,即 XA 必定被包含在 ρ 的一个子模式(R_i, ν)的 R_i 中,因此满足$\bigcup_{i=1}^{k} \pi_{(R_i, \mu_i)}(F) \mid = F$,由定义 16 得出,分解 ρ 保持 F 中的依赖。且算法在第一步中引入的额外的 TFD $R \rightarrow_{\mu} C$,将随着算法在第(4)步中对 C 属性的删除而不再成立。

(3)无损连接性

算法在第一步中引入了一个额外的 TFD $R \rightarrow_{\mu} C$,求规范覆盖的结果,该 TFD 左部消除了冗余属性后将保留下(R, μ)的一个时态关键字,因此,最后生成的 ρ 中必定有某个(R_i, μ_i),其中 R_i 包含此时态关键字,根据定理 7,可知 ρ 一定是无损连接的。

算法的时间复杂度。算法第(1)步使用的算法 TO-CANONICAL 的时间复杂度为 $O(p^2q^2)$,得到的规范覆盖集 G 中的 TFD 个数不会超过 pq,其中 q 为 R 中属性个数,p 为 F 中 TFD 的个数;算法第(2)步中第一个 for 循环是 $O(pqm)$ 级的,其中 m 为 F 中时态类型的个数,第二个 for 循环是 $O(pq^3m)$级的,故第(2)步的复杂度为 $O(pq^3m)$;算法第(3)步、第(4)步中的 for 循环是 $O(pq)$级的;算法第(5)步中的 for 循环是 $O(p^2q^2)$级的,因此,算法总的时间复杂为 $O(p^2q^2 + pq^3m)$。

结论 时态函数依赖是一种自然的时态约束。已经提出的基于多时间粒度时态函数依赖的范式分解算法相当复杂,在实际应用中难以用来进行有效的时态数据库设计。这种复杂性很大程度上是由于偏序时态类型间的复杂运算,而现实世界中的许多应用所涉及到的时态类型集都能满足全序关系,具有全序时态类型集的全序时态模块模式有着良好的特性,因此本文提出了全序时态模块模式、时刻关系模式、全序时态模块投影、全序时态简单候选关键字、全序时态三范式(TO_T3NF)等概念。由于 TO_T3NF 完全不能消除时态主属性对时态关键字的传递依赖,本文进一步提出了规范程度更高的全序时态简单范式(TO_TSNF)的概念,并给出了全序时态简单范式的分解算法,对其正确性及时间复杂度给予了证明与分析。下一步的主要工作是涉及时态多值依赖的全序时态模块模式的规范化研究。

参考文献

1 Segev A, Shoshani A. The Representation of a Temporal Data Model in the Relational Environment. In: Proceedings of the 4th International Conference on Statistical and Scientific Database Management, Rome, Italy, 1988. 39~61

2 Navathe S B, Ahmed R. A Temporal Relational Model and a Query Language. Information Sciences, 1989, 49:147~175

3 Jensen C S, Snodgrass R T, Soo M D. Extending existing dependency theory to temporal databases. IEEE Trans Knowledge and Data Engineering, 1996, 8(4):563~582

4 Wang X S, Bettini C, Jajodia S. Logical design for temporal databases with multiple granularities. ACM Trans Database System, 1997, 22(2):115~170

5 姚春龙,郝忠孝.具有全序时态类型集时态函数依赖集的研究[J].软件学报,2003,14(2):247~252

6 姚春龙,郝忠孝.一个多时间粒度下时态函数依赖的有限属性闭包算法.计算机研究与发展,2005, 42(3):448~454

7 Combi C, Rossato R. Temporal functional dependencies with multiple granularities: a logic based approach. In: 15th International Conference on Database and Expert Systems Applications (DEXA 2004), Zaragoza, Spain, August 30-September 3, 2004. 864~873

8 郝忠孝.关系数据库数据理论新进展.北京:机械工业出版社,1998