第 33 卷　第 9 期
Vol.33　*No.9*

计　算　机　工　程
Computer Engineering

2007 年 5 月
May 2007

·软件技术与数据库·　文章编号：1000—3428(2007)09—0076—03　文献标识码：A　中图分类号：TP311.13

FD环境下满足P_S及无α环模式分解问题研究

赵龄强[1]，郝忠孝[1,2,3]，顾照鹏[1]

(1. 哈尔滨理工大学计算机科学与技术学院，哈尔滨 150080；2. 齐齐哈尔大学计算机学院，齐齐哈尔 161006；
3. 哈尔滨工业大学计算机科学与技术学院，哈尔滨 150001)

摘　要：通过分析在FD集F的最小归并依赖集存在弱左部或弱右部冲突时所具有的性质和特征，讨论并给出了满足P_S(保持*FD*，无损连接且满足*SNF*)且无α环分解的充要条件和算法，对算法的正确性、可终止性进行了证明，并对算法的时间复杂度给出了分析。

关键词：无内部冲突；分解算法；无α环；简单范式

Research of Schema Decomposition Problem into Meeting P_S and Without α-cycle in Functional Dependency

ZHAO Lingqiang[1], HAO Zhaongxiao[1,2,3], GU Zhaopeng[1]

(1. College of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080; 2. College of Computer, Qiqihar University, Qiqihar 161006; 3. College of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

【Abstract】 By analyzing the property and characteristics of the minimum merge dependency set of F when D has weak left side conflict or weak right side conflict, the necessary and sufficient condition under which a database schema without inside conflict can be decomposed into one meeting P_S and α-acyclic is given. A decomposition algorithm is given accordingly with the proof for its termination and correction.

【Key words】 Without inside conflict; Decomposition algorithm, α-acyclic; Simple normal form

文献[2,3]分别讨论了*FD*集*F*无内部冲突情况下关系模式$R\langle W, F\rangle$完成了满足P_3、P_{BC}且无α环的分解问题。由于3*NF*的各种弊端，根据实际应用需要，在3*NF*的基础上对关键字进行约束，提出了略低于*BCNF*的*SNF*(简单范式)，见文献[1]。文献[4]讨论了数据库模式$R\langle W, F\rangle$中，*FD*集*F*无内部冲突，数据库模式分解成满足P_S且无α环的有关理论。本文在此基础上给出了数据库模式$R\langle W, F\rangle$，其*FD*集*F*无内部冲突条件下分解得到满足P_S且无α环的充要条件，并且给出了相应的模式分解算法以及正确性、可终止性证明，并对算法的时间复杂度进行了分析。

1 基本知识

定义 1[1] 设$R<U, F>$为一关系模式，U为属性集，F为函数依赖集。$X\subset U$，$A\in U$且$A\notin X$，$X\rightarrow A$。如果$R<U, F>$满足下列条件之一：

(1)X包含$R<U, F>$的一个简单候选关键字。

(2)存在一个简单候选关键字K，使$A\in K$。

则称$R<U, F>$满足简单范式，记为$R<U, F>\in SNF$。

定义 2[2] 对于一个*FD*集*F*，如果$F\vDash X\rightarrow Y$，$F\vDash Y\rightarrow X^{`}$均为非平凡的完全函数依赖，$X^{`}\subset X$，且$F\nvDash Y\rightarrow X$，*F*中存在内部冲突；否则，称*F*中无内部冲突。

定义 3[2] 一个*FD*集*F*的归并依赖集*D*定义如下：设*G*为*F*的一个最小覆盖，对*G*中左部等价的*FD*的左部属性集的集合作为*D*中的一个左部等价属性集的集合L_i，并把对应的*FD*的右部的并集中去掉可被L_i中某个左部属性集部分或传递函数决定的属性后所得到的属性集作为*D*中对应于L_i的右部RT_i。记W_{L_i}为L_i中所有属性的集合，$W_i=W_{L_i}\cup RT_i$，并使$RT_i=RT_i-W_{L_i}$，这样得到归并依赖集*D*为

$$D=\{d_i=L_i\rightarrow RT_i\}_{i\in N}$$

另用$X_i^1, X_i^2,\ldots$表示L_i中不同的等价左部属性集。

定义 4[2] 设*D*是*FD*集*F*的一个归并依赖集，对于*D*中的每一个$d_i=L_i\rightarrow RT_i$，若存在$j\neq i$，使$X_i^k\cap W_j\neq\varnothing(X_i^k\in L_i)$，则称$X_i^k$为对外相关的。如果通过等价属性集的替换，使*D*中每一个d_i中对外相关的左部等价属性集的个数最少，称替换后的结果集为最小归并依赖集。

定义 5[2] 设$D=\{d_i=L_i\rightarrow RT_i\}_{i\in N}$是*FD*集*F*的最小归并依赖集，$d_i$，$d_j\in D(i\neq j)$。如果$F\vDash X_i\rightarrow A$，$A\in W_{L_j}$，但$X_i\nrightarrow X_j$，且不存在属性集$X$，$A\in X$，使$X\leftrightarrow X_i$。则称$d_i$对$d_j$弱左部冲突。

定义 6[2] 设$D=\{d_i=L_i\rightarrow RT_i\}_{i\in N}$是*FD*集*F*的最小归并依赖集，$d_i$，$d_j\in D(i\neq j)$。如果$F\vDash X_i\rightarrow A$，$A\in RT_j$，且不存在属性集$X$，$A\in X$，使$X_i\leftrightarrow X$。同时，$F\nvDash X_i\rightarrow X_j$，$F\nvDash X_j\rightarrow X_i$则称$d_i$对$d_j$弱右部冲突。

定理 1[4] 设数据库模式$R\langle W, F\rangle$的*FD*集*F*无内部冲突，*D*为*F*的一个最小归并依赖集，如果*D*中存在弱左部冲突，则将$R\langle W, F\rangle$分解为满足条件P_S的数据库模式为有α环的。证明见文献[4]。

定理 2[4] 设数据库模式$R\langle W, F\rangle$的*FD*集*F*无内部冲突，*D*为*F*的一个最小归并依赖集，如果*D*中存在弱右部冲突，则将$R\langle W, F\rangle$分解为满足条件P_S的数据库模式为有α环的。证明见文献[4]。

定理 3[4] 设数据库模式$R\langle W, F\rangle$的*FD*集*F*无内部冲突，*D*为*F*的一个最小归并依赖集，如果*D*中不满足条件*T*，则将*R*

基金项目：黑龙江省自然科学基金资助项目(F00-06)

作者简介：赵龄强(1963－)，男，副教授、博士生，主研方向：关系数据库系统与理论；郝忠孝，教授、博导；顾照鹏，硕士生

收稿日期：2006-11-07　**E-mail**：zhaplq@hrbust.edu.cn

〈W，F〉分解为满足条件P_S的数据库模式为有α环的。证明见文献[4]。

定理 4[5] 如果数据库模式$R<W,F>$的一个分解$\{R_1, R_2, \ldots, R_n\}$是保持函数依赖的，且分解后有一个关系模式中含有候选关键字，则分解必是无损连接的。证明见文献[5]。

2 定义、定理

定义 7 关联度是描述最小归并依赖集D中每个L_i中的等价属性集与其它W_j相交非空的级别，用G_i表示。

(1)令$Q_i=W_{L_i}\cap(\bigcup_{j\neq i} W_j)$表示$d_i\in D$的左部对外相交的属性集。如果存在$d_i\in D$，$j\neq i$，$Q_i\subseteq Q_j$，且$Q_i\rightarrow W_i$，则令$G_i$:=-1。

对于这样的d_i必是可归约的。使$D=D-\{d_i\}$，再对D重复上面的操作，直至不再减少。

(2)对每个$d_i\in D$，令$Q_i=W_{Li}\cap(\bigcup_{di\in D} W_j)$，$j\neq i$，若Qi=□，则令Gi:=0。否则，若Qi□Xik，则Gi:=1。

(3)除(1)、(2)情况外，均令G_i:=2。

定义 8 设$Q=\{R_1,R_2,\ldots,R_n\}$是数据库模式R〈W，F〉的一个分解，若满足保持FD、无损连接性，且对于R_i中任一FD：$X\rightarrow A\in\prod_F(R_i)(XA\subseteq R_i, A\notin X)$，满足以下两个条件之一：

(1)X包含R_i的一个简单候选关键字；

(2)存在一个R_i的简单候选关键字K，使$A\in K$，则称该分解满足P_S。若分解除满足P_S外还具有无α环特性，称该分解满足P_S和无α环的分解。

定义 9 对FD集F和其最小归并依赖集$D=\{d_i=L_i\rightarrow RT_i\}_{i\in N}$，若存在$d_i\in D$的关联度 2，则与$d_i$相关联的$d_j$的关联度必为-1，称$D$满足$T$。否则称$D$不满足条件$T$。

定理 5 设关系数据库模式$R<W,F>$的FD集F，F是无内部冲突的。把$R<W,F>$可以分解得到同时满足保持函数依赖性、无损联结性、SNF及无α环 4 种特性的关系数据库模式分解的充要条件是FD集F的最小归并依赖集D中不存在弱左部冲突和弱右部冲突且是满足条件T的。

证明

必要性：由定理 1~定理 3 知，R〈W，F〉的任何具满足P_S且为无α环的分解，FD集F的最小归并依赖集是无弱左部、弱右部冲突且满足条件T的。

充分性：设D是FD集F的一个无弱左部冲突和弱右部冲突的最小归并依赖集。满足P_S的分解可能是多个，不一定都是无α环的。对D采用以下处理，能得到一个无α环的 分解。

(1)对 D 中的每个依赖，求其关联度，把关联度不为-1 的依赖并入集合 G 中；

(2)对G中的每个依赖d_i，建立一个属性集e_i，使e_i包含d_i的至少一个等价属性集。对d_i的左部中每个属于$W_{L_i}\cap(\bigcup_{j\neq i} W_j)$的属性$B(d\in G)$，均有$B\in e_i$，把$e_i$并入$G$中；

(3)对于任何一个$e_i\in G$，如果存在$d_j\in G(j\neq i)$，有$F\vDash X_j\rightarrow e_i$，则将$e_i$并入$D$中，并从$G$中删去$e_i$。

(4)将G中所有的e_i构成属性集V，把未出现在任何依赖中的属性并入V中；

(5)把每个$e_i\in D$作为一个关系模式，每个$d_i\in D$的W_i均作为一个关系模式，把属性集V构成一个关系模式；

(6)消去被其它关系模式所包含的关系模式。

该分解满足P_S和无α环特性。

(1)保持FD：每一个$d_i\in D$的W_i均作为一个关系模式，F与D具有等价性，因此得到的分解保持FD。

(2)无损连接性：对于G中的每个d_i，都有e_i与之对应，G中所有e_i的并集可以函数决定G中所有属性。步骤(4)把所有G中剩余的e_j并入属性集V，同时把所有外部属性并入到V中，故V是R〈W，F〉的一个超候选关键字，根据定理 4，分解是无损连接的。

(3)满足 SNF：

1)由每个$d_i\in D$所得的关系模式$R_i\in SNF$。d_i的右部RT_i不含传递或部分函数依赖于d_i的左部W_{L_i}属性，d_i的左部属性为R_i的主属性，W_{L_i}中的属性不存在部分依赖和传递依赖于不包含其候选关键字的属性，否则存在内部冲突，故$R_i\in SNF$。

2)由D中的每个e_j的关系模式$R_j\in SNF$。每一个e_i都是由d_i的左部的等价属性构成，D满足条件T，因此R_{ei}中每一个等价属性集均是R_{ei}的一个简单候选关键字，$R_{ei}\in SNF$。

3)由V得到的关系模式$R_v\in SNF$。由步骤(4)V是由外部属性和G中剩余的e_i构成的。由于最小归并依赖集D是无弱左部冲突的，因此G中剩余的每一对e_{ij}之间不存在任何依赖关系，V所产生的关系模式$R_v\in SNF$。

只有这 3 种途径产生关系模式，故该分解满足 SNF。

(4)无α环性：用$Graham$算法对ρ对应超图处理。关联度为-1 的归并依赖构成的模式先后被约去，剩下包含由D中e_i、G中的d_j和V构成的关系模式。D不存在弱左部和弱右部冲突，设由d_i构成的关系模式R_{d_i}，有$W_{N'}(R_{d_i})\cap RT_i=\varnothing$，于是$W_{N'}(R_{d_i})\subseteq W_{L_i}$。由步骤(2)有$W_{N'}(R_{d_i})\subseteq e_i$，则$R_{dj}$可被归约掉。$e_i$构成的关系模式$R_{ei}$，若不存在$d_j$且$R_{dj}$未被归约，使$W_{L_j}\cap e_i\neq\varnothing$，则必存在$d_k$且$RT_k\cap e_i\neq\varnothing$。由$D$不存在弱左部冲突，于是$X_k\rightarrow e_i$。又由于不存在弱右部冲突，因此不存在$d_m$，$RT_m\cap e_i\neq\varnothing$，则有$W_{N'}(R_{ei})\subseteq R_{dj}$，$R_{ei}$可以被归约。按这样的规则，由$D$中的$e_i$和$G$中的$d_j$产生的关系模式先后归约去。最后剩下由$V$得来的关系模式，亦可归约。即$Graham$算法成功，因此分解结果无$\alpha$环。

3 算法、时间复杂度

算法 1 GLD (求最小归并依赖集D中每个d_i的关联度)

输入：归并依赖集 D；

输出：每个d_i的关联度G_i；

begin
  call Mini-GB(D)[6] //求最小归并依赖集
  for 每一个 $d_i\in D$ *do* //求d_i的对外相交属性集
    $Qi=W_{Li}(\bigcup_{j\neq i} W_j)$;
  for 每一个 $di\in D$ *do*
    if $Q_i\subseteq X_i^m$ *and* $Q_i\subseteq W_j$ *then* G_i=:-1; //求出关联度为 1 的d_i
  for 每一个 $d_i\in D$ *do*
    if $G_i\neq$-1 *then if* $Q_i=\varnothing$ *then* G_i:=0;
    //除去关联度为-1 计算其它d_i的关联度
    else if $Q\subseteq X_i^k$ *then* G_i:=1;
   else G_i:=2;
 return(G_i)
end

由关联度的定义，算法是正确的。第(1)步时间复杂度为$O(P^3)$，第(2)、第(3)步时间复杂度分别为$O(P)$、$O(P^2)$，因此算法 1 时间复杂度为$O(P^3)$，其中P为输入的FD的个数。

算法 2 $T\_Test(G)$(测试最小归并依赖集 G 是否满足条件 T 的算法)

输入：最小归并依赖集 G；

输出：若 G 满足条件 T 则输出 F，若 D 不满足条件 T 则

输出 T；

begin

$GLD(G)$; //求关联度

for 每一个 $d_i \in G$且$G_i=2$ *do*

for 每一个 $d_j \in G$，$(j \neq i)$ *do*

if $W_{Li} \cap W_{Lj} \neq \varnothing$ *and* $G_j \neq -1$ *then* $return(T)$;

$return(F)$

end

根据条件T的定义，知算法正确。属性个数为n，FD个数为P，算法第(1)步时间复杂度为$O(P^3)$，第(2)步最坏情况与P^2同级，故该算法时间复杂度为$O(P^2(P+1))$。

算法 3 FJ{F无内部冲突，分解为满足P_S且无α环的分解}

输入：$R<W, F>$；

输出：$R<W, F>$的一个分解；

begin

$D:=GB\_Set(F)$[6] //求F的归并依赖集

$G:=Mini\text{-}GB(D)$[6] //求D的最小归并依赖集G

if $C\text{-}test(G, F)$[7] *or* $T\_Test(G)$ *then* $Cyclic(G)$[7]，$return(\rho)$;

//D 不同时满足无弱左、右部冲突且条件 T，返回有 α 环分解

$GLD(D)$; //求关联度

$G(0):=\varnothing$;

for 每一个$d_i \in G$ *do*

if $G_i \neq -1$ *then* $G(0):=G(0) \cup \{d_i\}$;

//把关联度不为-1 的d_i并入$G(0)$中//

for 每一个$d_i \in G(0)$ *do* $e_i:=\varnothing$;

for 每一个$X_i^k \in L_i$ *do* $Q_l^k:=\varnothing$;

for 每一个$d_j \in G(0)$且$j \neq i$ *do*

$Q_l^k:=Q_l^k \cup (X_l^k \cap W_j)$; //求每个$X_l^k$的对外相交属性集

for 每一个$X_l^k \in d_i$ *do*

if $Q_l^k \neq \varnothing$且不存在$Q_l^k \subseteq Q_i^n (n \neq k)$ *then* $e_i:=e_i \cup X_l^k$;

//考查每个X_l^k是否属于e_i//

if $e_i=\varnothing$ *then* $e_i:=min(L_i)$

//d_i中无左部属性对外相交，则取L_i中长度最小的//

$G(0):=G(0) \cup \{e_i\}$;

for 每一个$e_i \in G(0)$ *do*

if 存在 $d_j \in G(0)(i \neq j)$ 且 $(W_j \cap e_i) \rightarrow e_i$ *then* $G:=G \cup \{e_i\}$; $G(0):=G(0)-\{e_i\}$; //消去被其它X_j所包含的e_i

$K:=\varnothing$;

for 每一个$e_i \in G(0)$ *do* $K:=K \cup e_i$;

$K:=K \cup (W - \cup W_j)$; //外部属性加入候选关键字中

for 每一个d_i，$e_j \in G$ *do*

把W_i或e_j作为一个关系模式;

把 K 作为一个关系模式;

for 每一个$R_i \in \rho$ *do*

if 存在$R_i \subseteq R_j (i \neq j)$ *then* $\rho:=\rho-\{R_i\}$;

//消去被其它R_j所包含的R_i

$return(\rho)$

end

显然根据定理 5，算法FJ是正确的。属性个数为n，FD个数为P，算法(1)、(2)、(3)、(4)步的时间复杂度分别为$O(nP)$、$O(P^3)$、$O(P(P^2+P))$、$O(P^3)$，其它各步均与P同级或小于P，故该算法的时间复杂度为$O(P(P^2+P+n))$。

4 结束语

本文给出了在FD环境下F无内部冲突时，分解产生满足条件P_S的数据库模式的充要条件，并且给出了具体的分解算法、正确性和时间复杂度分析。

参考文献

1 郝忠孝. 关系数据库理论新进展[M]. 北京: 机械工业出版社, 1998.

2 郝忠孝. 无内部冲突满足P_3的无α环的数据库模式分解(I): 分解的基本理论[J]. 计算机研究与发展, 1998, 35(4): 301-304.

3 邓成玉. 关联度与关系模式分解满足P_{BC}且无α环的关系[J]. 计算机工程与应用, 2004, 40(18): 181-182, 232.

4 赵龄强. 数据库模式在FD环境下满足P_S及无α环判定问题研究[J]. 计算机工程, 2007, 33(1).

5 郝忠孝. 连接超图的有关理论研究(II): 无 α 环分解的基本理论[J]. 计算机研究与发展, 1997, 34(增刊): 263-266.

6 郝忠孝. 最小归并依赖集求解算法[J]. 计算机研究与发展, 1997, 34(增刊): 267-270.

7 郝忠孝. 无内部冲突满足P_3的无α环的数据库模式分解(II): 分解的算法及分析[J]. 计算机研究与发展, 1998, 35(4): 305-309.

~~~~~~

（上接第 75 页）

对 $y \in H^{-}X$，则 $H(y) \cap X \neq \varnothing$，由性质 1，$H(y) \subseteq L(y)$，从而 $L(y) \cap X \neq \varnothing$，于是 $y \in L^{-}X$，即 $H^{-}X \subseteq L^{-}X$。

由文献[5]及性质 2，得 $T\_X \subseteq L\_X \subseteq H\_X$，$H^{-}X \subseteq L^{-}X \subseteq T^{-}X$。因此，权衡容差关系是比容差关系和限制容差关系都要优越的一个更好的相似关系。

3 结论

针对带缺省属性值的不完备信息系统，本文提出一种权衡容差关系。通过分析得出，该相似关系比容差关系、非对称相似关系、限制容差关系和修正容差关系等相似关系更加合理有效，也更具有实用价值。

参考文献

1 张桂芸, 黄国兴, 杨炳儒. 基于分辨相似矩阵的相似粗糙集的属性约简算法[J]. 计算机工程, 2006, 32(10): 43-44, 65.

2 Pawlak Z, Busse J G. Rough Sets[J]. Communications of the ACM, 1995, 38(11): 89-95.

3 Kryszkiewicz M. Rough Set Approach to Incomplete Information System[J]. Information Sciences, 1998, 11(2): 39-49.

4 Stefanowski J, Tsoukias A. On the Extension of Rough Sets Under Incomplete Information[C]//Zhong S, Skowron A, Ohsuga S. Proc. of the 7th Int'l Workshop on New Directions in Rough Sets, Data Mining, and Granular Soft Computing. Berlin: Springer-Verlag, 1999: 73-81.

5 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238-1243.

6 刘富春. 基于修正容差关系的扩充粗糙集模型[J]. 计算机工程, 2005, 31(24): 145-147.

7 刘富春. 不完备信息系统中粗糙集理论的扩充[J]. 计算机与现代化, 2005, (6): 1-3, 7.

8 林泰崴, 陈昭炯, 叶东毅. 不完备信息系统下限制容差关系模型的改进[J]. 福州大学学报(自然科学版), 2004, 32(增刊): 39-42.
~~~~~~