

基于线图的无 β 环混合依赖分解条件及算法*

刘文远^{1,2}, 郝忠孝^{1,3}

¹哈尔滨工业大学 计算机科学与工程系, 黑龙江 哈尔滨 150001)

²(燕山大学 计算机科学与技术系, 河北 秦皇岛 066004)

³哈尔滨理工大学 计算机系, 黑龙江 哈尔滨 150080)

E-mail: wylu@ysu.edu.cn

摘要: 无 β 环数据库模式具有很多优良的特性, 以往的研究都局限在图论的范畴内, 而没有考虑数据库的其他规范化特性. 在混合依赖概念的基础上, 定义了严格无冲突、扩展严格无冲突等概念, 并证明了在混合环境下得出的无损联接、保持依赖、无 β 环且满足4NF的分解的充要条件是, 混合依赖集是扩展严格无冲突的. 据此, 给出了判断严格无冲突及混合环境下无 β 环分解算法, 并分析了算法时间的复杂度是线性的. 最后, 给出基于线图的实例验证. 这一结论可直接指导数据库的模式设计.

关键词: 数据库模式; 混合依赖; 线图; 无 β 环; 无损联接; 保持依赖

中图法分类号: TP311 文献标识码: A

在各类信息系统(MIS, MRPII, ERP等)建立的过程中, 数据库技术起着非常重要的作用. 人们希望自己的数据库中的数据模式能够正确描述现实世界各种信息之间的内在联系. 在这方面, Codd和Armstrong等人提出了一套完整的规范化理论. 为了减少数据库设计中存在的异常现象, 20世纪80年代, C. Beeri等人又提出了无环数据库模式^[1], 无环成为判断数据库模式优劣的另一重要特性. 在对无环的研究过程中发现, 数据库模式存在着不同级别的无环性, R. Fagin在文献[2]中描述了3种重要的无环类型, 即无 α 环、无 β 环和无 γ 环, 其中对无 α 环的讨论比较深入^[1,3,4]. 但无 α 环有一个缺陷, 就是无 α 环超图的子图可能仍是有环的, 这同样可能导致查询错误. 无 β 环除了具有无 α 环的所有优良特性以外, 它还有一个重要特性, 即其每个子图都是无环的. 这使得无环级别更高.

在数据库模式设计中, 有很多重要的特性, 如无损联接、保持依赖等需要考虑, 而且必须满足某一范式的要求, 只有综合考虑上述条件并满足无环性, 才能设计出一个优秀的数据库模式. 本文的研究重点是在混合依赖环境下, 讨论得出无损联接、保持依赖、无 β 环且满足4NF的分解的条件和性质, 并给出了分解算法. 这一结论能够直接指导具体数据库的设计, 避免数据库设计中因出现无环等问题而导致的失败.

1 无损联接、保持依赖、无 β 环且满足4NF的分解条件

定义 1.1. 设 Γ 是一个混合依赖集, X 为 T 上的左部属性集, 设 $DEP(X) = \{A_1, \dots, A_m \parallel Y_1, \dots, Y_k\}$ 为 Γ 的依赖基. 称 M_x 为属性集 X 的简并MVD(multivalued dependency)集, 其中 $M_x = \{X' \twoheadrightarrow Y_1, \dots, X' \twoheadrightarrow Y_k\}$, $X' = X \cup \{A_1, \dots, A_m\}$.

若 $X' = U$ (即 $K=0$ 时), M_x 定义为 $\{U \twoheadrightarrow \emptyset\}$. 若 M 是 Γ 中每个左部属性集的简并MVD集的集合, 则MVD集 M 是 Γ 的简并MVD集.

当存在一个FD(function dependency)和MVD的混合依赖集时, 文献[5]中证明了一个简并MVD集可以完全表达对混合依赖集的全部语义信息. 这样就可以用MVD集代替一个混合依赖集来解决分解等问题.

* 收稿日期: 1999-05-18; 修改日期: 1999-09-20

作者简介: 刘文远(1968-), 男, 黑龙江密山人, 讲师, 主要研究领域为, 数据库理论与应用, ERP研究; 郝忠孝(1940-), 男, 山东蓬莱人, 教授, 博士生导师, 主要研究领域为关系数据库理论, 无环数据库理论.

定义 1.2. 设 M 是属性集 R 上的一个 MVD 集, $LHS(M)$ 为 M 中所有 MVD 左部属性集的集合, $X, Y, Z \in LHS(M)$, 若 M 满足以下条件, 则称 M 是严格无冲突的:

- (1) M 中每一个 MVD 都不分裂其他 MVD 的左部;
- (2) 对于 M 中的任意两个 MVD 的左部属性集 X 和 Y , $(DEPM(X) \cap DEP_M(Y)) \subseteq DEP_M(X \cap Y)$;
- (3) $X, Y \subseteq LHS(M)$, $X_i \in DEP(X)$, $Y_j \in DEP(Y)$, 设 $A = \{a_1, \dots, a_m \mid a_i \in Y\}$, $B = \{b_1, \dots, b_n \mid b_i \in Y_j\}$, $AB \neq Y$ 且 $AB \subseteq LHS(M)$. 若 $Z = XX_i \cap YY_j$, 则 Z 包含于 AB .

定义 1.3. 设 Γ 是一个混合依赖集, 若 Γ 中的简并 MVD 集是无冲突的, 则 Γ 是扩展无冲突的; 若简并 MVD 集是严格无冲突的, 则称 Γ 是扩展严格无冲突的.

定理 1.1. 在纯 MVD 环境下, 产生满足无损联接、保持依赖、无 β 环且满足 4NF 的分解的数据库模式的充要条件是给定的 MVD 集存在一个严格无冲突的覆盖.

证明: 略.

定理 1.2. 设 Γ 是一个混合依赖集, M 是 Γ 的简并 MVD 集, 设 G 是 Γ 中所有 FDS 和 M 中所有 MDS 的集合, 则 $\Gamma = G$.

证明: 因 G 包含 Γ 中所有 FDS, 故 Γ 中和 G 中的 FDS 是相等的, 下面只需证明简并 MVD 集与原 Γ 中 MVDS 集是等价的即可. 假设 X 是左部属性集, 根据文献[5], X 在 Γ 中的依赖基 $DEP(X)$ 和 X 在 M 中的依赖基 $Dep(X)$ 是相等的, 也就是说, 任意 $Y \in DEP(X)$, 则 $Y \in Dep(X)$, 反之亦然. 因此, M 中的 MVDS 与 Γ 中的 MVDS 是相等的, 故 $\Gamma = G$. \square

定理 1.3. 设 Γ 是一个混合依赖集, M 是 Γ 的简并 MVD 集, 设 D 是 Γ 中所有 FDS 的集合, 则 D 中每个 FD 是与 M 相容的.

证明: 根据简并 MVD 集的定义, X 是 Γ 的左部属性集, 当 X 既是 FD 也是 MVD 的左部时, 即 $DEP(X) = \{A_1, \dots, A_m \parallel Y_1, \dots, Y_k\}$, $m \neq 0, k \neq 0$. 这时, $X^* = X \cup \{A_1, \dots, A_m\}$, 故每个 FD $X \rightarrow A_i (1 \leq i \leq m)$ 都包含于 X^* , 故是相容的. 若 X 只是 FD 的左部时, 即 $DEP(X) = \{A_1, \dots, A_m\}$, 也就是说, $m \neq 0, k = 0$. 根据定义有 $X^* = U$, M_x 定义为 $U \rightarrow \emptyset$, 这时, 任何 FD $X \rightarrow A_i (1 \leq i \leq m)$ 都包含于 U , 当然是与 M 相容的. \square

定理 1.4. 设 Γ 是一个混合依赖集, M 是 Γ 的简并 MVD 集, 把 Γ 分解为满足无损联接、保持依赖、无 β 环且满足 4NF 的分解的充要条件为, 是扩展严格无冲突的.

证明: (1) 充分性: 设 G 是 Γ 中所有 FDS 和 M 中所有 MVDS 的集合, 根据扩展严格无冲突的定义, 当 Γ 是扩展严格无冲突的, 说明 Γ 的简并 MVD 集是扩展严格无冲突的, 根据定理 1.2, G 和 Γ 是等价的, 根据定理 1.3, G 中每个 FD 都和 G 中的简并 MVD 集是相容的. 也就是说, 简并 MVD 集中的 MVDS 包含所有 FD 的语义信息, 因简并 MVD 集 M 是严格无冲突的, 根据定理 1.1, 分解一定是无损联接、保持依赖且满足 4NF 的.

(2) 必要性: 根据定理 1.1, 若给定的 MVD 集可得到一个无 β 环且满足无损联接、保持依赖、无 β 环的分解, 则此 MVD 集是严格无冲突的. 因此, 若 Γ 的简并 MVD 集 M 可得到满足无损联接、保持依赖、无 β 环且满足 4NF 的分解, 则简并 MVD 集 M 是严格无冲突的, 根据定义 1.3, Γ 一定是扩展严格无冲突的. \square

定义 1.4. 数据库模式 $R = (R_1, R_2, \dots, R_n)$ 的线图 $LG(R) = (V, E, L)$ 包括图 $G(R) = (V, E)$ 和一组标记 L , 定义如下:

- (1) 当且仅当 $R_i \in R; v_i \in V, v_i$ 的标记 $L(v_i)$ 是属性集 R_i .
- (2) 当且仅当 $v_i, v_j \in V$ 且 $L[e] \neq \emptyset, e = (v_i, v_j)$, 其中 $L[e] = L[v_i] \cap L[v_j]$ 称为边 e 的权.

定义 1.5. 当且仅当 $U' = (v'_1, e'_1, v'_2, e'_2, \dots, v'_m, e'_m, v'_1)$ 是 $LG(R')$ 的纯环时, G 中的环 $U = (v_1, e_1, v_2, e_2, \dots, v_m, e_m, v_1)$ 称为 $LG(R) = (V, E, L)$ 的 β 环, 其中 $R' = \{R'_i - R_i - X (\text{空 } R'_i \text{ 则删除})\}_{i=1, \dots, m}, v'_i = \{L[v_i] - X\}, e'_i = \{L[e_i] - X\}, 1 \leq i \leq m$, 且 $X = L[v_1] \cap \dots \cap L[v_m]$.

若一个线图满足是三角化的且其中任一三角形是不相容的, 则称其为无 β 环的^[6].

2 无损联接、保持依赖、无 β 环且满足 4NF 的分解算法

算法 2.1. STRICT-CONFLICT-FREE (判断 MVD 集 M 是否为严格无冲突的)

输入: MVD 集 M 和泛属性集 U ;

输出: 若 M 是严格无冲突的, 输出 true, 否则输出 false.

STRICT-CONFLICT-FREE(M, U)

(1) If CONFLICT FREE(M') = false then return (false);

(2) $X, Y \subseteq LHS(M')$

for 每个 $X_i \in DEP(X)$ do

for 每个 $Y_j \in DEP(Y)$ do

$A = \{a_1, \dots, a_m \mid a_i \in Y_i\}, B = \{b_1, \dots, b_n \mid b_i \in Y_j\}, AB \neq Y$ 且 $AB \subseteq LHS(M)$

$Z = XX_i \cap YY_j$, If Z 不包含于 AB then return (false)

(3) return (true)

算法 2.2. FM- β -ACYCLIC-DECOMP(混合环境下无 β 环分解算法)

输入: 一个混合依赖集 Γ 及泛属性集 U ;

输出: 一个无 β 环的满足 P_4 的数据库模式 R .

FM- β -ACYCLIC-DECOMP(Γ, U)

(1) M -DEGENERATED-MVD(Γ)^[4];

(2) If STRICT-CONFLICT-FREE(M) = false then return (false);

(3) $Q = \{X_1, \dots, X_k\}, X_i \in LHS(M)$ 且对任何 $X_i \leq X_j$, 有 $1 \leq i \leq j \leq k$;

(4) $R =$ DECOMP(M, Q, U);

(5) Return (R).

3 算法分析及应用

若 MVD 中属性个数为 m , MVD 个数为 n . 算法 DEGENERATED-MVD 是求混合依赖集 Γ 的简并 MVD 集 M , 算法时间复杂度为 $O(mn)$.

在算法 STRICT-CONFLICT-FREE 中, 算法 CONFLICT-FREE 判断是否是无冲突的, 时间复杂度为 $O(nm^4 + n^2)$, 第(2)步的时间复杂度为 $O(m^4 + n^2)$, 故算法总的时间复杂度为 $O(nm^4 + n^4)$.

在算法 FM- β -ACYCLIC-DECOMP 中, 第(4)步是 $O(mn^2)$, 故算法总的时间复杂度是 $O(nm^4 + n^4)$.

例: 设混合依赖集 $\Gamma = \{C \twoheadrightarrow F, CE \twoheadrightarrow G, CE \twoheadrightarrow I\}, U = CEFGHI, \Gamma$ 的简并 MVD 集 $M = \{C \twoheadrightarrow F, CEI \twoheadrightarrow C\}$, 因此, M 的左部偏序序列为 $Q = \{C, CEI\}$, 首先用 C 对 U 进行分解, 得到 CF 和 $CEGHI$, 再用 CEI 对 $CEGHI$ 进行分解, 得到 $\rho = \{CF, CEIG, CEIH\}$, 根据定义 1.4, 构建 ρ 的线图, 可知此线图是三角形且是相容的, 故容易验证 ρ 是满足无损联接、保持依赖、无 β 环且满足 4NF 的分解.

4 结论

本文提出了严格无冲突的概念. 在混合依赖环境下, 在求出混合依赖集对应的简并 MVD 集, 并证明了 MVD 集是严格无冲突时, 也就是在混合依赖集是扩展严格无冲突时, 可以分解得到无损联接、保持依赖、无 β 环且满足 4NF 的数据库模式, 并给出了相应的算法, 最后用一个实例进行验证.

References:

[1] Beeti, C., Fagin, R., Maier, D., et al. On the desirability of acyclic database schemes. Journal of ACM, 1983, 30(3): 479~513.

[2] Fagin, R. Degrees of acyclicity for hypergraphs and relational database schemes. Journal of ACM, 1983, 30(3): 514~550.

[3] Hao Zhong-xiao, Yao Chun-long, Gao Yan. Research on the theoried concerning the join hypergraph(I): basic concepts. Computer Research and Development, 1997, 34(supplement): 259~262 (in Chinese).

- [4] Hao Zhong-xiao, Gao Yan, Yao Chun-long. Research on the theoried concerning the join hypergraph(II): basic theory on α -acyclic hypergraph decompositions. Computer Research and Development, 1997, 34 (supplement): 263 ~ 270 (in Chinese).
- [5] Katsuno, H. An extension of conflict-free multivalued dependency sets. ACM Transactions on Database System, 1984, 9(2): 309 ~ 326.
- [6] Cheung, T. Y., Zhu, Y. Z. Recognizing different types of beta-cycles in a database scheme. Theoretical Computer Science, 1991, 81(3): 295 ~ 304.

附中文参考文献:

- [3] 郝忠孝,姚春龙,高岩. 连接超图的有关理论研究(I):基本概念. 计算机研究与发展, 1997, 34(增刊): 259 ~ 262.
- [4] 郝忠孝,高岩,姚春龙. 连接超图的有关理论研究(II):无 α 环分解的基本理论. 计算机研究与发展, 1997, 34(增刊): 263 ~ 270.

Decomposition Condition and Algorithm of β -Acyclic in Mixed Dependency Environments Based on Line Graph

LIU Wen-yuan^{1,2} HAO Zhong-xiao^{1,3}

¹(Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001, China);

²(Department of Computer Science and Technology, Yanshan University, Qinhuangdao 086004, China);

³(Department of Computer, Harbin University of Science and Technology, Harbin 150080, China)

E-mail: wyluu@ysu.edu.cn

Received May 18, 1999; accepted September 20, 1999

Abstract: β -Acyclic database scheme has a number of desirable properties, but most former researches about β -acyclic are based on graph theory, without consideration on other canonical properties of database. With the concept of mixed dependency basis, this paper defines the concepts of strict conflict-free and extended strict conflict-free, and proves that decomposition satisfies lossless join, keeps dependency, β -acyclic and 4NF if and only if it is extended strict conflict-free in the mixed environment. Based on this, the paper gives the algorithm for judging strict conflict-free and β -acyclic decomposition in mixed environment, then shows that the complexity of the algorithm is linear using an example based on line graph. This conclusion can directly guide real database designing.

Key words: database scheme; mixed dependency; line graph; β -acyclic; lossless join; keep dependency