第5卷第9期 软件学报 Vol. 5, No. 9
1994年9月 JOURNAL OF SOFTWARE Sept. 1994

具有无损联结性且保持依赖性关系模式的BCNF完备分解算法*

徐庆生 周行仁

（云南大学计算机科学系，昆明 650091）

摘要 本文给出一种具有完备性的合成方法，来把一个关系模式分解成具有无损性和保持依赖性的BC范式关系数据库模式，只要这个模式"本质上能作这种分解的话"；同时对这种"本质上能分解为保持某些性质的某一范式"提法进行了形式化描述；最后，讨论了这种合成法的固有复杂度.

关键词 函数依赖，分解，无损联结，保持函数依赖，BC范式.

关系数据库规范化过程的一个重要问题是如何保持信息完整性和完整性约束，即如何使关系模式到某一范式的分解具有无损联结性和保持函数依赖性.当前研究要么是在某一特殊子集上进行[1,2]，要么是针对一些程度较低的范式进行[3].本文试图从理论上提出关系模式的具有无损性和保持函数依赖性的BCNF分解方法.首先，介绍一种函数依赖推导图概念，然后，讨论分解方法以及分解方法的复杂度.文中基本概念或术语见文献[4—7].

1 有向无环图(DAG)

定义1.1. 设关系模式 $R(U,F)$，其中 U 是属性集，F 是函数依赖集.基于 F 的推导有向无环图是一个用 U 中属性符号标定的有向无环图，简称基于 F 的 $DDAG$，或 $F-DDAG$，可根据下列规则构造：

$R1$. 任何一个带有 U 中符号的不相联结点集都是一个 $F-DDAG$.

$R2$. 设 H 是一个 $F-DDAG$，它包括分别带有符号 $A_1,A_2,\cdots,A_k$ 的结点 $v_1,v_2,\cdots,v_k$. 再设 $A_1,A_2,\cdots,A_k\to CZ\in F$. 通过把一个带符号 C 的结点 u 和边 $\langle v_1,u\rangle,\langle v_2,u\rangle,\cdots,\langle v_k,u\rangle$ 加入到 H 中形成 H'. H' 是一个 $F-DDAG$.

$R3$. 除此之外都不是 $F-DDAG$.

定义1.2. 设 H 是一个 $F-DDAG$，v 是 H 中一个结点.如果 v 没有任何进入的有向边，则称 v 是 H 的初始结点.

---

* 本文1991—08—08收到，1992—05—30定稿

作者**徐庆生**，31岁，助教，主要研究领域为数据库，程序设计方法和面向对象.**周行仁**，59岁，副教授，主要研究领域为程序设计方法学，数据库技术.

本文通讯联系人：徐庆生，楚雄 675000，楚雄师范专科学校计算中心

定义 1.3. 设 H 是一个 $F-DDAG$. 如果下列条件成立，则称 H 是 $X\to Y$ 的基于 F 的 $DDAG$：

$D1$. X 是初始结点的符号集；且 $D2$. Y 中每一个属性都标识了 H 中某一结点.

定义 1.4. 已知 H 是一个 $F-DDAG$. 定义 H 的有用集是在构造 $DDAG$ 过程利用 $R2$ 规则使用了的 F 中依赖组成，记为 $u(H)$. 如果 H 是 $X\to Y$ 的 $F-DDAG$，则有用集也记为 $uF(X\to Y)$ 或 $u(X\to Y)$.

引理 1.1.[4]　设关系模式 $R(U,F)$，$X\to Y$ 是一个函数依赖. 下列给论是等价的：

(1)F 逻辑推导出 $X\to Y$；(2)存在一个关于 $X\to Y$ 的 $F-DDAG$.

引理 1.2.[4]　如果 H 是 $X\to Y$ 的 $F-DDAG$ 且 $V\to W\in u(H)$，则 F 推导出 $X\to V$.

2　基本分解算法

本文用 $A,B,\cdots$ 表示单属性，$X,Y,\cdots$ 表示属性集.

定义 2.1. 设关系模式 $R(U,F)$ 和关系数据库模式 $\overline{R}=(R1(U1,F1),\cdots,Rn(Un,Fn))$. 如果 $\bigvee_{i=1}^{n} Ui=U$，则称 $\overline{R}$ 是 R 的一个分解. 再设 r 是 $R(U,F)$ 的任一关系. 如果 $R=\Pi R2(r)*\Pi R2(r)*\cdots*\Pi Rn(r)$，则称 $\overline{R}$ 关于 R 具有无损性；如果 $F\equiv\bigvee_{i=1}^{n} Fi$，则称 $\overline{R}$ 关于 R 或 F 具有保持依赖性.

定义 2.2. 设 F 是最小函数依赖集，$X\to A\in F$. 如果下列条件成立，则称 $X\to A$ 满足 BC 模式条件或简称 $X\to A$ 满足 BCC：对 X 的任一划分 $\{X1,X2\}$，其中 $X1$ 可为空集且 $X2\neq\Phi$，如果存在单属性 $B\in X2$ 使得 F 推出 $X1A\to B$，则 F 推导出 $X1A\to X2$. 如果 F 中每一依赖都满足 BCC，则称 F 满足 BCC.

定理 2.1. 设关系模式 $R(U,F)$，其中 U 是属性集，F 是最小函数依赖集. 对 F 作一划分 $\{F',F'',F'''\}$，其中 $F'=\{X\to A\mid X\to A\in F$ 且 X 关于 F 不是 R 的码$\}$，$F''=\{X\to A\mid X\to A\in F-F'$ 且存在 $Y\to B\in F'$ 使 $B\in XA\}$，$F'''=F-F'-F''$. 这里 F'，F'' 和 F''' 不会同为空. 对每一依赖 $X\to A\in F'\vee F''$ 都设计一关系模式，其属性集为 XA 且对集合 F''' 设计一关系模式，其属性集由 F''' 中所有依赖涉及到的属性组成. 把这些模式记为 $\overline{R}=(R1(U1,F1),\cdots,Rn(Un,Fn))$，其中 $Fi\equiv\pi Ri(F)$，$1\leqslant i\leqslant n$. 则有

(1)$\overline{R}$ 关于 F 保持函数依赖.

(2)若 $F'\vee F''$ 满足 BCC，则 $\overline{R}\in BCNF$；反之亦然.

(3)若 $F''\vee F'''\neq\Phi$，则 $\overline{R}$ 具有无损性.

(4)若 $\overline{R}$ 中加入一关系模式 $R0$，其属性集 X 是 R 的码，则扩充的 $\overline{R}$ 有无损性且 $R0\in BCNF$.

在证明定理 2.1 前先给出一些预备性质. 因篇幅有限这里略去.

引理 2.1. 设 F 是最小函数依赖集，且 $X\to A\in F$. 如果存在 XA 的非空子集 $X1$ 和 $X-X1$ 的子集 B 使 F 推出 $X1\to B$，则 $A\in X1$.

引理 2.2. 设 $R(U,F)$，F 是最小依赖集. 如果 F 满足 BCC，则对任何 $X\to A\in F$，所设计的关系模式 $R'(U',F')$(其中 $U'=XA$，$F'\equiv\pi R'(F)$)属于 $BCNF$.

引理 2.3. 令 F，F'，F''' 与定理 2.1 相同. 若 $F'''\neq\Phi$，则关系模式 $R0(U0,F0)\in BCNF$，

其中 $U0$ 由 F''' 涉及到的属性组成，$F0 \equiv \pi R0(F)$.

引理 2.4. 设 X 是 $R(U,F)$ 的一个码，$R'(U',F')$ 是一关系模式，其中 F 是最小依赖集，$U'=X$，$F' \equiv \pi R'(F)$. 则 $R' \in BCNF$.

引理 2.5. 设 $\overline{R}$ 是 R 的一个保持依赖的分解. 如果存在 $R' \in \overline{R}$ 且 R' 的属性集 U' 是 R 的超码，则 $\overline{R}$ 关于 R 具有无损性[4,5].

由引理 2.1—2.5 可证明定理 2.1 如下：

(1)由假设知，对 $X \to A \in F$，存在 $i \in \{1,2,\cdots,n\}$ 使 $Ri \in \overline{R}$ 且 Ui 包含 XA，所以 $X \to A \in Fi$.

从而 $\bigvee_{i=1}^{n} Fi$ 推出 $X \to A$，于是 $\bigvee_{i=1}^{n} Fi$ 推出 F，即 $\overline{R}$ 具有保持依赖性.

(2)设 $F' \vee F''$ 满足 BCC 且 $R' \in \overline{R}$，则 R' 由 $F' \vee F''$ 中一依赖构成或由 F''' 构成，据引理 2.2 和 2.3，$R' \in BCNF$，所以 $\overline{R} \in BCNF$. 反之，设 $\overline{R} \in BCNF$，$X \to A \in F' \vee F''$，则存在 $R' \in \overline{R}$ 使得 R' 的属性集 U' 包含 XA，且对 X 的任何划分 $X1$ 和 $X2 \neq \Phi$，只要存在 $B \in X$ 使 F 推出 $X1A \to B$，据假设 $R' \in BCNF$，那么 F 推出 $X1A \to X2$，所以 $X \to A$ 关于 F 满足 BCC. 故 $F' \vee F''$ 满足 BCC.

(3)如果 $F'' \vee F''' \neq \Phi$，则存在 $Ri(Ui,Fi) \in \overline{R}$，它是由 F'' 中某一依赖或者由 F''' 构成，所以 Ui 是 R 的一个超码，据引理 2.5，$\overline{R}$ 关于 R 具有无损性.

(4)如果 $\overline{R}$ 中加入一关系模式 $R0(X,F0)$，其中 X 是 R 的码. 则由引理 2.5 知 $\overline{R} \vee \{R0\}$ 具有无损性. 又由引理 2.4 知 $R0 \in BCNF$.

根据定理 2.1 可得到下列具有无损性且保持依赖性的 $BCNF$ 的基本分解算法.

算法 2.1. 具有无损性且保持函数依赖性 $BCNF$ 分解.

输入：最小依赖集 F 及其涉及到的属性集 U.

输出：若 F 满足 BCC，则返回具有无损性且保持依赖性 $BCNF$ 数据库模式 $\overline{R}$；否则"φ".

$LLPFBC(F,U)$

(因篇幅有限，细节省略，由定理 2.1 不难得出)

3 分解算法的完备性

定义 3.1. 函数依赖族定义为如下集合，将其记为 FDC：$\{F \mid F$ 是一个最小函数依赖集$\}$.

定义 3.2. 设 $F \in FDC$. 如果存在 F 的一个分解 $\overline{R}(R1(U1,F1),\cdots,Rn(Un,Fn))$ 使得 $F \equiv \bigvee_{i=1}^{n} Fi$，则称 F(或 $R(U,F)$)本质上可分解为保持依赖的数据库模式. 如果 $\overline{R} \in BCNF$，则称 F 本质上可分解为保持依赖的 $BCNF$ 数据库模式.

定义 3.3. 定义 $FDC1=\{F \mid F \in FDC$ 且本质上可分解为保持依赖的 $BCNF$ 数据库模式$\}$，$FDC2=FDC-FDC1$.

引理 3.1. 设关系模式 $R(U,F)$，其中 F 是最小依赖集，即 $F \in FDC$. 如果 $F \in FDC1$，则 $R(U,F)$ 存在一个最小依赖集 F' 满足 BCC.

定义 3.4. 设 $X \to A \in F$，这里 F 是最小依赖集. 定义 $USBCC(X \to A)=\{Y \to B \mid F$ 推出

$X \to Y, Y \to X$ 且非平凡的 $Y \to B$ 满足 $BCC\}$ 以及 $USBCC(F) = \bigvee_{X \to A \in F} USBCC(X \to A)$.

引理 3.2. 设 F 和 F' 是等价的两个最小依赖集. 则对任何 $X' \to A' \in F'$ 存在 $X \to A \in F$ 使得 F 推出 $X' \to X, X \to X'$.

其证明可借助引理 1.1 和 1.2 得出，这里因篇幅有限略去.

引理 3.3. 已知最小函数依赖集 F. 从 $USBCC(F)$ 出发通过左既约和消除冗余依赖所得到的 $USBCC(F)$ 的最小依赖集满足 BCC.

引理 3.4. 如果存在 $X \to A \in F$ 使得 $USBCC(X \to A) = \Phi$，则 $USBCC(F)$ 不等价于 F.

定理 3.1. 设 F 是最小依赖集. $USBCC(F) \equiv F$ 当且仅当 $F \in FDC1$.

证明：设 $USBCC(F) \equiv F$，用 G 记 $USBCC(F)$ 按照引理 3.3 求得的最小依赖集，则 G 满足 BCC. 根据定理 2.1，G 有一分解 $\overline{R} \in BCNF$ 且 $\overline{R}$ 关于 G 具有保持依赖性. 但 $G \equiv USBCC(F) \equiv F$，所以 $\overline{R}$ 关于 F 具有保持依赖性. 故 $F \in FDC1$.

反之，设 $F \in FDC1$. 则由引理 3.1，存在与 F 等价的最小依赖集 G 满足 BCC. 再由引理 3.2，对任何 $X \to A \in G$，有 $Y \to B \in F$ 使 F 推出 $X \to Y, Y \to X$，但 $X \to A$ 满足 BCC，所以 $X \to A \in USBCC(Y \to B)$，从而 $X \to A \in USBCC(F)$. 于是 $USBCC(F)$ 推出 G，即 $USBCC(F)$ 推出 F. 又由定义 3.4，F 推出 $USBCC(F)$，所以 $USBCC(F) \equiv F$.

算法 3.1. 扩充的具有无损性且保持依赖性 $BCNF$ 分解.

输入：最小依赖集 F 及其涉及到的属性集.

输出：若 $F \in FDC1$，则返回具有无损性且保持依赖性的 $BCNF$ 数据库模式；否则返回"Φ".

$EXP-LLPFBC(F, U)$

(1) $\overline{R} := LLPFBC(F, U)$;

(2) 如果 $\overline{R} \neq \Phi$ 则返回 $\overline{R}$；否则继续执行；

(3) 求出 $USBCC(F)$ 的最小依赖集 G；

(4) 如果 G 不等价于 F，则返回 Φ；否则转步(5)；

(5) 返回 $LLPFBC(G, U)$.

4　分解算法时间复杂度分析

算法 2.1 和算法 3.1 的时间复杂度分析可归结为下述问题.

问题 1. 实例：属性集 U，U 上最小依赖集 F，U 的子集 Z. 问：Z 关于 F 是否违反 BC 范式？

引理 4.1. 问题 1. 是 NP 完全的[8].

问题 2. 实例：属性集 U，U 上最小依赖集 F，依赖 $X \to A \in F$. 问：XA 关于 F 是否违反 BC 范式？

定理 4.1 问题 2. 是 NP 完全的.

证明：我们可以在多项式时间内判定 XA 上任一合法的非平凡依赖的决定因素是否是 XA 的码，所以问题 2 是 NP 问题. 设问题 1. 的任一实例：U, F, Z，其中 Z 包含于 U 中. 由该实例可以在多项式时间内构造问题 2 的一个实例：$U' = UD$（其中特殊的单属性 D 不属于

$U)$，$F'=\{WD\to B \mid W\to B\in F\}\vee\{Z\to D\}$，$X'=Z$，$A'=D$. 这里 F' 是最小依赖集，证明如下：如果有 $WD\to B\in F'$ 使 $F'-\{WD\to B\}$ 推出 $WD\to B$，则 $F'-\{WD\to B,Z\to D\}$ 推出 $WD\to B$，所以 $F-\{W\to B\}$ 推出 $W\to B$，矛盾. 如果有 $WD\to B\in F'$ 且 $C\in W$ 使 $(WD-\{C\})\to B$ 能由 F' 推出，则 F 推出 $(W-\{C\})\to B$，矛盾；又 $Z\to D$ 不是 F' 的冗余依赖，所以 F' 是最小依赖集.

现在证明 $X'A'$ 关于 F' 违反 BC 范式当且仅当 Z 关于 F 违反 BC 范式. 当 Z 关于 F 违反 BC 范式，由 F' 的定义知，ZD 关于 F' 违反 BC 范式；当 ZD 关于 F' 违反 BC 范式，由引理 2.1，存在 Z 的一个划分 $Z1$，$Z2$ 和 $Z3$ 使得 F' 推出 $Z1D\to Z2$ 且 $Z3$ 不依赖于 $Z1D$，于是 F 推出 $Z1\to Z2$ 且 $Z3$ 不依赖于 $Z1$，所以 Z 关于 F 违反 BC 范式. 至此已证问题 2 是 NP 完全的.

参考文献

1 Yuan L Y *et al*. Logical design of relational database schemes. Proc. of the 6th ACM PODS, 1987:38－34.

2 Yuan L Y *et al*. Unifying functional database and multivalued dependencies for relational database design. Proc. of the 5th ACM PODS, 1986(3):183－190.

3 崔亨洙. 简单关键字和简单范式. 计算机学报, 1985, 7(4):262－267.

4 Maier D. The theory of relational database, 1983.

5 Ullman J. D. Principles of database systems. 2nd ed., 1983.

6 萨师煊, 王珊. 数据库系统概论. 北京: 高等教育出版社, 1991.

7 施伯乐等. 数据库理论及新领域. 北京: 高等教育出版社, 1990.

8 Garey M R, Johnson D S, 张立昂(译). 计算机难解性. 1987.

THE BCNF NORMALIZATION METHOD OF RELATIONAL DATABASE SCHEME: HAVING A LOSSLESS JOIN AND PRESERVATION OF DEPENDENCIES

Xu Qingsheng and Zhou Xingren

(*Department of Computer Science, Yunnan University, Kunming* 650091)

Abstract This paper presents a complete synthetic approach to decompose a relational scheme into a database scheme which has a lossless join and preservation of dependencies and which is in BC normal form so long as the relational scheme essentially having this decomposition nature, and it also provides a formal description of the concept——essentially having this decomposition nature. This paper also analyses the time complexity of the synthetic approach.

Key words Functional dependency, decomposition, lossless join, dependencies－preservation, BC normal form.