第7卷第4期
1996年4月

软 件 学 报
JOURNAL OF SOFTWARE

Vol. 7, No. 4
Apr. 1996

外部闭包及其在关系数据库中的应用

王宇君　施伯乐　胡美琛

(复旦大学计算机系　上海　200433)

摘要　本文引入了广义FD路和外部闭包的概念,将它们应用于函数依赖集的无冗余覆盖计算,有效地减少了计算的闭包个数.并在此基础上提出了一个新的3NF合成算法,将常用的3NF合成算法中的2次无冗余覆盖计算合并为1次,显著地减少了计算闭包总个数.

关键词　广义FD路,外部闭包,无冗余覆盖,3NF合成算法.

关系数据库RDB(relational database)理论的核心:规范化理论在70年代末已基本建立,但对它的研究一直十分活跃.如文献[1]中发现,通过对3NF,BCNF数据库模式添加一个简单条件,可以保证4NF,5NF的自动获取.文献[2]给出了一个新的多项式时间的BCNF RDB的设计算法.文献[3]阐述了常数时间可维护BCNF数据模式的概念、测试算法和有关性质.其中G. Ausiello等[4]首先提出了FD图的概念,指出用FD图来表示函数依赖(以下简称FD),闭包、覆盖、最小化等问题可得到类似的处理,并给出了效率更高的算法.D. Jim等[5]提出了一种新的闭包:$r-$闭包,能减少求无冗余覆盖过程中计算的闭包个数,并考虑了将$r-$闭包应用至3NF合成算法,以减少闭包的计算个数.文献[6]对文献[4]中FD图理论作了进一步的发展和完善.文献[7]中提出了一种有向超图,并将它应用至RDB领域,给出了对FD等概念的很自然的解释.

基于广义FD路,本文提出了外部闭包的概念,它是一种图结点闭包,又吸取了$r-$闭包的思想,如应用于FD集的无冗余覆盖计算,它和$r-$闭包一样,在消除结点的冗余虚(实)弧时,能减少不必要的闭包的计算.同时,由于外部闭包包含的信息比$r-$闭包多,前者是图结点的集合,后者是属性的集合,因此,本文得出了一些较文献[5]更强的结论,如对等价左边的判断等.并给出了计算外部闭包的线性时间的算法.且只须对初始化步骤稍作修改,该算法还可以用来计算结点闭包和结点虚闭包.

文献[5,8]中的3NF合成算法对于输入的FD集,先计算一次无冗余覆盖,接下来为了保证分解的模式个数最少,需要合并等价左边,而合并等价左边可能带来新的冗余,这样,就需要再计算一次无冗余覆盖.这在各种文献、教科书中已成为共识(参见本文附录中的例3).本文将无冗余覆盖计算合二为一,并给出了详细的证明.再加上外部闭包的应用,显著地

---

* 作者王宇君,1970年生,博士生,主要研究领域为数据库理论及应用.施伯乐,1935年生,教授,博士导师,主要研究领域为知识库,面向对象数据库及数据库新领域.胡美琛,女,1938年生,副教授,主要研究领域为数据库理论及应用.

本文通讯联系人:王宇君,上海200433,复旦大学计算机系

本文1995-01-14收到修改稿

减少了计算闭包的总个数.因而本文提出的3NF算法较以往3NF算法,效率有较大改善.

1 基本概念

定义 1. 设R(U)是一关系模式,X,Y⊆U,R上存在**函数依赖** FD:X→Y是指,对R的任一实例r中的任意元组t_1,t_2,若$t_1[X]=t_2[X]$,则$t_1[Y]=t_2[Y]$.

若一FD右端为单属性,则称该FD是**简单的**.由这样的FD所构成的集合被称为**简单FD集**.

定义 2. 设F是给定属性集U上的非平凡简单FD集,它对应的FD**图**是标号图$G_F=\langle N_F,A_f,A_d\rangle$,其中:

$N_F=N_s\cup N_c$是结点集,N_s是简单结点集,N_c是复合结点集.对U中每个单属性A在G_F中构造标号A的简单结点;对F中每一条FD:X→A且|X|>1在G_F中构造标号X的复合结点*.

A_f是实弧集.对F中每一条FD:X→A,在G_F中标号为X,A的结点间构造有向实弧(X,A).

A_d是虚弧集.对F中每一条FD:X→B,若X不是单属性,则对每一个A∈X,在G_F中标号为X,A的结点间构造有向虚弧(X,A).

定义 3. 对FD图$G_F=\langle N_F,A_f,A_d\rangle$,定义:

A的**源**为X,若$(X,A)\in A_d$.A的**源集** $S(A)=\{X|(X,A)\in A_d\}$.

X的**邻接集** $AS(X)=\{A|(X,A)\in A_f\cup A_d\}$.即$X$的邻接结点集.

X的**虚邻接集** $DAS(X)=\{A|(X,A)\in A_d\}$.

X的**实邻接集** $FAS(X)=\{A|(X,A)\in A_f\}$.

X的**邻接内集** $IAS(X)=\{A|A\in FAS(X)$且结点A的实弧出度=0且虚弧入度=0$\}$.

X的**邻接外集** $EAS(X)=\{A|A\in AS(X)$且结点A的实弧出度>0或虚弧入度>0$\}$.

显然$AS(X)=IAS(X)\cup EAS(X)$,且$IAS(X)\cap EAS(X)=\varnothing$.

定义 4. 对FD图$G_F=\langle N_F,A_f,A_d\rangle$,从结点集$\Sigma$到结点$Y$的**广义$FD$路**$\langle\Sigma,Y\rangle$是满足下述任一条件的$G_F$的最小子图$\overline{G}_F=\langle\overline{N}_F,\overline{A}_f,\overline{A}_d\rangle$:

①(邻接规则)$\overline{A}_f\cup\overline{A}_d=\{(Z,Y)|Z\in\Sigma,Y\in AS(Z)\}$.

②(实弧传递规则)Y是简单结点,存在结点Z满足$Y\in FAS(Z)$,且$\overline{G}_F$中有广义FD路$\langle\Sigma,Z\rangle$.

③(虚弧并规则)Y是复合结点,$Y\notin\Sigma$,且对任意$A\in DAS(Y)$,有或者在$\overline{G}_F$中存在广义FD路$\langle\Sigma,A\rangle$,或者$A\in\Sigma$.

$\overline{G}_F$是最小子图意味着在$\overline{G}_F$中再删去任意的弧或结点后,广义FD路$\langle\Sigma,Y\rangle$要求的3个条件都不成立.

若广义FD路$\langle\Sigma,Y\rangle$满足离开结点集Σ的所有的弧都是虚弧,则称该FD路是广义FD虚路,否则称之为广义FD实路.

---

* 此时FD中的X代表一个由2个或2个以上的属性所组成的属性集,而G_F中的X代表图中一个结点.

这里的广义 FD 路是文献[4]中的 FD 路的推广,但以下为表达简洁仍简称 FD 路. 当 Σ 中仅包含一个结点 X 时,我们习惯将 FD 路 $\langle\{X\},Y\rangle$ 简记为 $\langle X,Y\rangle$.

定义 5. 对 FD 图 $G_F=\langle N_F,A_f,A_d\rangle$,定义:

X 的**虚闭包**为:$X_d^+=\{Y\mid G_F$ 中存在 FD 虚路 $\langle X,Y\rangle\}$.

X 的**闭包**为:$X^+=\{Y\mid G_F$ 中存在 FD 路 $\langle X,Y\rangle\}$.

X 的**实闭包**为:$X_f^+=X^+-X_d^+$.

G_F 的**闭包**为标号图 $G_F=\langle N_F,A_f^+,A_d^+\rangle$,其中:

①$(X,A)\in A_d^+$ 当且仅当在 G_F 中 $A\in X_d^+$;

②$(X,A)\in A_f^+$ 当且仅当在 G_F 中 $A\in X_f^+$.

定义 6. 对 FD 图 G_F 中,若存在一条不包括虚(实)弧 (X,Y) 的 FD 虚(虚或实)路 $\langle X,Y\rangle$,则**称虚(实)弧 (X,Y) 是冗余的**.

在 FD 图 G_F 中,若对每一条形如 (X,A) 的实弧,其中 X 为复合结点,都存在一条 FD 虚路 $\langle X,A\rangle$,则**称复合结点 X 是冗余的**.

例 1:设 U=ABCDE,F={A→D,AB→C,AB→E,E→D,AD→C,BD→E},则其对应的 FD 图如图 1(a)所示,其中,$S(A)=\{AB,AD\}$,$DAS(AB)=\{A,B\}$,$FAS(AB)=\{C,E\}$,$IAS(AB)=\{C\}$,$EAS(AB)=\{A,B,E\}$,$AB^+=\{A,B,C,D,E,AD,BD\}$,$AB_d^+=\{A,B,C,D,E,AD,BD\}$.

FD 实路 $\langle AB,D\rangle$,FD 虚路 $\langle AB,BD\rangle$ 分别如图 1(b)和图 1(c)所示. 由图 1(a)中可看出,虚弧 (AD,D) 冗余,实弧 (AB,C) 冗余,复合结点 AB 冗余.

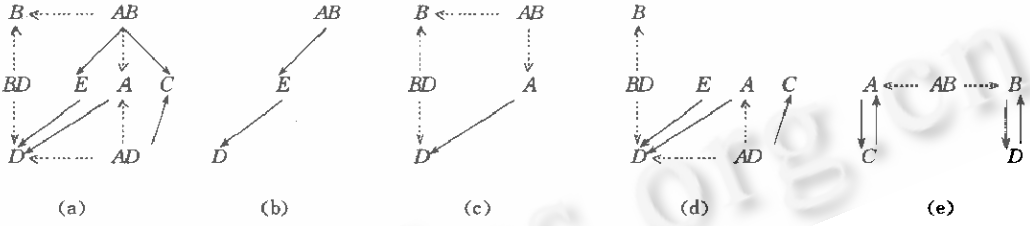

图1

2　外部闭包

定义 7. 对 FD 图 $G_F=\langle N_F,A_f,A_d\rangle$,设 $G^*(X)$ 为 G_F 中删去从 X 发出的所有弧(包括实弧和虚弧)后得到的 FD 图,定义 X 的**外部闭包**为:

$X^e=\{Y\mid FD$ 图 $G^*(X)$ 中存在 FD 路 $\langle EAS(X),Y\rangle\}$.

例 2:考虑图 1(a)中的 FD 图,$G^*(AB)$ 为图 1(d)所示,由图可知 $AB^e=\{Y\mid\langle\{A,B,E\},Y\rangle\}=\{C,D,E,AD,BD\}$.

对 X 的闭包 X^+ 的计算可以看成分为如下 2 步:

①$X^+\leftarrow AS(X)$.

②$X^+\leftarrow X^+\cup X^e$.(在 X 的邻接集全部加到 X^+ 中后,结点 X 发出的弧对闭包的计算不会有新的贡献,可以从图中删去. 又邻接内集中的结点也不会有新的贡献. 故只需考虑从邻接外集出发可达的结点集. 外部闭包由此而命名).

由外部闭包的定义可得：

定理 1. $X^+=X^e\cup AS(X)$.

定理 2. 简单结点 $A\in X^e$ 当且仅当存在结点 $Y, Y\neq X$, $A\in FAS(Y)$ 且 $Y\in X^+$.

证明：⇐，此时结点 Y 和弧 (Y,A) 显然都在 $G^*(X)$ 中，再由 $Y\in X^+$ 知 $Y\in EAS(X)$ 或 $G^*(X)$ 中存在 FD 路 $\langle EAS(X),Y\rangle$，故 $G^*(X)$ 中 FD 路 $\langle EAS(X),A\rangle$ 存在，即 $A\in X^e$.

⇒，$A\in X^e$，则 FD 图 $G^*(X)$ 中存在 FD 路 $\langle EAS(X),A\rangle$，设该路中最后一条弧为 (Y,A)，由 FD 路定义及 $EAS(X)$ 中无复合结点可知，(Y,A) 为实弧且 $Y\neq X$. 对 Y 有 $Y\in EAS(X)$ 或者 FD 路 $\langle EAS(X),Y\rangle$ 存在. 所以，在原 G_F 中存在 FD 路 $\langle X,Y\rangle$ 有 $Y\neq X$, $A\in FAS(Y)$ 且 $Y\in X^+$. □

定理 3. 设 (X,A) 是 $G=\langle N,A_f,A_d\rangle$ 中的冗余弧，则 $A\in X^e$.

证明：若弧 (X,A) 冗余，则图 G_F 中存在不包含 (X,A) 的 FD 路 $\langle X,A\rangle$，设该路中最后的弧为 (Y,A)，显然 $Y\neq X$（否则与 $\langle X,A\rangle$ 不包含 (X,A) 矛盾），易知 (Y,A) 为实弧，由定理 2 得 $A\in X^e$. □

定理 4. 设 (X,A) 是 $G_F=\langle N_F,A_f,A_d\rangle$ 中的实弧，若 $A\in X^e\cap IAS(X)$，则 (X,A) 冗余.

证明：若 $A\in X^e$，则图 $G^*(X)$ 中存在 FD 路 $\langle EAS(X),A\rangle$，设图 G_F 中对应的 FD 路为 $\langle X,A\rangle$，因为 $A\in IAS(X)$，所以 $\langle X,A\rangle$ 中不含弧 (X,A)，故 (X,A) 冗余. □

定理 5. 对已消除了冗余虚弧的 FD 图 $G_F=\langle N_F,A_f,A_d\rangle$，①简单结点 X 存在等价结点的充要条件是 $X\in X^e$；②复合结点 X 有等价结点的必要条件是 $DAS(X)\subseteq X^e$.

证明：①若简单结点 X 有等价结点，设为 A, $A\neq X$，则 G_F 中存在 FD 路 $\langle X,A\rangle$, $\langle A,X\rangle$，设 $\langle A,X\rangle$ 中最后一条弧为 (Y,X)，易知 (Y,X) 为实弧，且 $Y\neq X$, $Y\in X^+$，应用定理 2 可得 $X\in X^e$.

若 $X\in X^e$，则在 $G^*(X)$ 中存在 FD 路 $\langle EAS(X),X\rangle$，设该 FD 的最后一条弧为 (Y,X)，显然 $Y\neq X$. 易知该弧为实弧，且或者 $Y\in EAS(X)$，或者 $G^*(X)$ 中 FD 路 $\langle EAS(X),Y\rangle$ 存在，2 种情况都有，在 G_F 中 FD 路 $\langle X,Y\rangle$ 存在，故 X 与 Y 等价.

②若复合结点 X 有等价结点，设为 Z, $Z\neq X$，则对任意 A, $A\in DAS(X)$，有 G_F 中存在 FD 路 $\langle Z,A\rangle$，该路也一定在 $G^*(X)$ 中. 设该路的最后一条弧为 (Y,A)，易知它为实弧且 $Y\neq X$，由定理 2 知，$A\in X^e$. 由 A 的任意性可知 $DAS(X)\subseteq X^e$. □

注意，当结点为复合结点时，必要条件不能改为充要条件，反例见图 1(e). 图中有 $DAS(AB)\subseteq AB^e$，但 AB 没有等价结点.

算法 1.（求结点外部闭包）

输入：FD 图 $G_F=(N_F,A_f,A_d)$，结点 X.

输出：X 的闭包 $CLO(X)$.

步骤：1. 初始化令 $T\leftarrow EAS(X)$, $CLO\leftarrow\varnothing$

2. 对每个复合结点 i，若 X 是简单结点且 $i\in S(X)$，则 $q[i]=1$，否则 $q[i]=0$

3. 当 $T\neq\varnothing$，执行：

3.1 从 T 中选择结点 j，并令 $T\leftarrow T-\{j\}$

3.2 若 j 是简单结点，则对每一个 $m\in S(j)-\{X\}$，执行：

3.3 $q[m]=q[m]+1$

3.4 若 $q[m]=|DAS(m)|$ 则 $T\leftarrow T\cup\{m\}$, $CLO\leftarrow CLO\cup\{m\}$

3.5 对结点 k,$k\in FAS(j)$,且 $j\neq X$,若 $k\notin CLO$,则 $T\leftarrow T\cup\{k\}$,$CLO\leftarrow CLO\cup\{k\}$

4. 输出 CLO.

对算法 1 稍作修改,就可用来求结点闭包和结点虚闭包. 在步骤 1 中,若令 $T\leftarrow EAS(X)$,$CLO\leftarrow AS(X)$,则输出的 CLO 为结点闭包;若令 $T\leftarrow DAS(X)$,$CLO\leftarrow T$ 则输出的 CLO 为结点虚闭包. 读者可自行验证.

3　求无冗余覆盖的快速算法

定义 8. 对 FD 图 G 及 G',若 $G^+=G'^+$,则称 G' 是 G 的**覆盖**.

若 G' 中既不包含冗余虚(实)弧,又不包含冗余结点,则称 G' 是 G 的**无冗余覆盖**.

算法 2. (求无冗余覆盖的快速算法)

输入:FD 图 $G_F=\langle N_F,A_f,A_d\rangle$.

输出:图 G_F 的无冗余覆盖 G_F'.

步骤:1. 对图 G_F 的每一个结点 X,若 $EAS(X)\neq\varnothing$

1.1 计算 X^e

1.2 若 $|DAS(X)|>1$,则对每一个 $B\in DAS(X)\cap X^e$,执行

1.3 在算法 1 的步骤 1 中令 $T\leftarrow DAS(X)-\{B\}$,$CLO\leftarrow\varnothing$,再执行步骤 2,3,4,设输出为 CLO

1.4 若 $B\in CLO$ 中,则删去虚弧 (X,B) 并相应修改 FD 图

2. 对每一条实弧 $(X,A)\in A_f$,若 $A\in X^e$,则执行

2.1 若 $A\in IAS(X)$,则删去 (X,A)

2.2 否则,在删去 (X,A) 后计算 X^+,若 $A\notin X^+$,则恢复 (X,A)

3. 对每一个复合结点 X,若 $FAS(X)=\varnothing$,则删去 X

4. 输出图 G_F.

算法中步骤 1 删去冗余虚弧,步骤 2 删去冗余实弧,步骤 3 删去冗余结点,步骤 1 和 2 的次序不能颠倒. 否则反例见附录例 4,并见文献[9]中第 436 页上的练习 7.13.

算法在步骤 1 中对 B 的选择和 2 中对 A 的选择,其正确性由定理 3、定理 4 保证. 到步骤 3 时,由于步骤 2 中已删去了所有的冗余实弧,故这时结点冗余当且仅当结点只有虚弧.

算法 2 对邻接外集非空的结点增加了一次外部闭包的计算,但对每一条满足 $A\notin X^e$ 的弧 (X,A) 或者 $A\in X^e\cap IAS(X)$ 的实弧 (X,A),省去了相应的闭包的计算,在输入为面向实际的 FD 图时,这 2 种冗余是很多的[5],因而算法 2 将大大减少闭包的计算个数.

由于算法 2 中计算 X^+ 只是为了测试某结点 A 是否属于 X^+,因此,实践中还可改变算法 1 的返回条件而得到一个判别算法,这样平均每次对 X^+ 的计算只有原来的一半.

4　3NF 合成的快速算法

以下用有限集对 $\langle R(U),F\rangle$ 表示 RDB,其中 $R=\{R_1(X_1),R_2(X_2),\ldots\}$ 是关系集,$F=\{F_1,F_2,\ldots\}$ 是 FD 集. 设 S_1,S_2 是 2 个 RDB 簇,关系数据库设计可以看成从 S_1 到 S_2 的映照:$f=\langle f_r,f_c\rangle$,对任意 $\langle R,F\rangle\in S_1$,有 $f(\langle R,F\rangle)=\langle f_r(\langle R,F\rangle),f_c(\langle R,F\rangle)\rangle$.

3NF 设计可以看成从 $\langle R(U),F_1\rangle$ 到 $\langle\rho(R),F_2\rangle$ 的转换过程,其中 $\rho(R)$ 中每一个关系模式都是 3NF 的.

算法 3. (3NF 合成的快速算法)

输入:$\langle R(U),F_1\rangle$

输出:$\langle \rho(R),F_2\rangle$

步骤:1. 初始化

1.1 令 $U_1=\{A|$属性 A 在 F_1 中出现过$\}$,$U_2=U-U_1$

1.2 任取没有在 U 中出现过的某虚属性 γ,令 $F_1=F_1\cup\{U_1\rightarrow\gamma\}$

1.3 建立相应的 FD 图 $G_{F_1}=\langle N_{F_1},A_f,A_d\rangle$

2. 删去 G_{F_1} 中的冗余虚弧(算法 2 的步骤 1)

并建立连通分量链 $G_1,G_2,\cdots,G_k$,初始化令每个 G_i 为恰包含一出度>0 的结点

3. 增加标识弧,对每一个结点 X,若有 $X\in X^e$,或 $DAS(X)\subseteq X^e$,则:

3.1 对每一个 $Y\in X^e$,若 $X\in Y^e\cup AS(Y)$,则:

3.2 若 Y 为简单结点,考虑结点对(X,Y);否则依次考虑所有形如(X,A),$A\in DAS(Y)$的结点对:若结点对之间有弧相连且为实弧,则标识该弧;若无任何弧相连则相应增加一条实弧并标识它.

3.3 合并 X 和 Y 所在的连通分量链 G_i,G_j

4. 删去 G_F 中的冗余实弧

4.1 对步骤 3 中没有标识的弧,执行算法 2 的步骤 2

4.2 对步骤 3 中标识过的弧,执行算法 2 的步骤 2

5. 删去冗余结点(算法 2 的步骤 3)

6. 构造关系模式

6.1 设虚结点 γ 所属的连通分量链为 G_i,在 G_i 中删去该虚结点

6.2 令 $\rho\leftarrow\{ATT(i)\cup U_2\}$

6.3 对其它每一个连通分量 $G_j(j\neq i)$,令 $\rho\leftarrow\rho\cup\{ATT(j)\}$

6.4 令 F_2 为化简后的 G_F 对应的 FD 集,输出$\langle\rho(R),F_2\rangle$.

步骤 3 中对结点的选择的正确性由定理 5 保证,步骤 3.1 中的 $Y^e\cup AS(Y)$ 即 Y^+. 具体实现时标识弧的方法可选择在弧上添加标志位,或者简单地将指针指向的结点号取负. 步骤 6 中的 $ATT(j)$为 G_j 中出现过结点的邻接集对应的属性的集合. 如在附录例 3 中,当算法 3 执行至步骤 6 时,设 $G_1=\{AB,AD\}$, $G_2=\{C,D\}$,此时 $ATT(1)=\{A,B,D\}$, $ATT(2)=\{C,D\}$.

该算法只求了一次无冗余覆盖,而 3NF 算法[5,8]需计算 2 次,因此该算法减少了闭包个数的计算,再加上消去冗余弧及判断等价结点时外部闭包的应用,算法 3 与以往 3NF 算法相比,时间复杂性 $O(k*n^6)$中系数 k 得到了显著减小.

下面考虑算法的正确性. 文献[10]中指出 Bernstein 在提出 3NF 合成算法时,其基本思想是:对于给定的一个泛关系模式 $R(U,F)$,根据 F 进行合成,得到一个数据库模式,$\rho=\{R_1,R_2,\ldots,R_k\}$,使之具有下述特征:

①$\rho\in$3NF;

②ρ 保持 F,并且 $F\equiv\{K_i\rightarrow R_i|i\in\{1,2,\cdots,n\}\}$;

③对于 R 上任何满足 F 的泛关系实例 r,$r=r_1\infty r_2\infty\cdots\infty r_k$;

④ρ 在所有满足上述条件的数据库模式中所包含的关系模式个数最少.

其中,性质②简称 F 的(键)约束可表示性,性质③保证模式的无损连接性,性质④可减少数据冗余. 我们证明算法 3 满足这 4 个性质.

引理 1. 添加标识弧不会带来新的虚弧冗余.

证明:用反证法. 设图 $G_F=\langle N_F,A_f,A_d\rangle$,添加实弧$(Y,B)$后,导致虚弧$(X,A)$冗余,设

添加(Y,B)后的 FD 图为 G'_F,即 G'_F 中存在不包括(X,A)的虚路$\langle X,A\rangle$,则(Y,B)一定出现在该路中,否则 G_F 中也包含该 FD 路$\langle X,A\rangle$,于是(X,A)在 G_F 中就会冗余,得出矛盾.

又因为 $B\in Y^+$,即 G_F 中有 FD 路$\langle Y,B\rangle$.该 FD 路中一定包括弧(X,A),否则我们用该 FD 路取代 G'_F 中的$\langle X,A\rangle$中的弧(Y,B),并作路的最小化,就得到了不包括(X,A)的虚路$\langle X,A\rangle$,即$\langle X,A\rangle$在 G_F 中也是冗余的,与题设矛盾.

于是在 G'_F 中,由(Y,B)一定出现在 FD 路$\langle X,A\rangle$中可知,G'_F 中存在 FD 路$\langle X,Y\rangle$,该路既不包括弧(Y,B),又不包括弧(X,A),故该路也在 G_F 中;另外,在 G_F 中,存在一定包括弧(X,A)的 FD 路$\langle Y,B\rangle$,则 G_F 中存在不包括(Y,B)的 FD 路$\langle Y,A\rangle$,且由(X,A)是虚弧可知,$\langle Y,A\rangle$中也不包括(X,A).2 种结合起来,并作路的最小化,可得 G_F 存在 FD 路$\langle X,A\rangle$,它不包括(X,A),而这与假设(X,A)在 G_F 中不冗余矛盾.□

定理 6. 算法 3 步骤 2～5 得到图 G_F 的无冗余覆盖.

证明:由引理 1 及算法 2 即可得该结论.□

由定理 6 及与文献[8]中的 3NF 算法比较,我们可以得到定理 7.

定理 7. 算法 3 得到的分解满足性质①是 3NF 的,②是(键)约束可表达的.

定理 8. 算法 3 得到的分解满足性质④,即关系模式的个数最少.

证明:由于我们对等价左边进行了合并,由文献[10]中引理 10.3 及其推论即得证.

定理 9. 算法 3 得到的分解满足性质③.

证明:由于我们在算法 3 的初始化步骤中添加了函数依赖 $U_1\rightarrow Y$,并且在步骤 6 中将 F_1 中没有出现过的属性集 U_2 添加到了含 U_1 中的关键字的连通分量链 G_i 中,故 G_i 中含有原泛关系 U 的键,这样就保证了算法 3 的无损连接性.□

参考文献

1 Date C J, Fagin R. Simple conditions for guaranteeing higher normal forms in relational databases. ACM Trans. Database Syst., 1992,**17(3)**:465～476.

2 Zhang Y C, Orlowska M E. A new polynomial time algorithm for BCNF relational database design. Information Systems, 1992,**17(2)**:185～193.

3 Hernandez H J, Chan E P F. Constant－time－maintainable BCNF database schemes. ACM Trans. Database Syst., 1991,**16(4)**:571～599.

4 Ausiello A G, D'Atri, Sacca D. Graph algorithms for functional dependency manipulation. J. ACM30 1983, 1983. 752～766.

5 Diederich Jim, Milton Jack. New methods and fast algorithms for database normalization. ACM Trans. Database Syst., 1988,**13**:339～365.

6 Ausiello G, D'Atri A, Sacca D. Minimal representation of directed hypergraphs. SIAMJ. Comput., 1986,**15**:418～431.

7 Gallo G, Longo G, Pallottino S. Directed hypergraphs and applications. Discrete Applied Mathematics, 1993,**42**:177～201.

8 Yao S Bing ed. Principles of database design. Volume1: Logical Organizations, Prentice－Hall Inc., Englewood Cliffs, 1985.

9 Ullman J D. Principles of database and knowledge－base systems. Volume I, New York: Computer Science Press, 1990.

10 施伯乐,何继潮,崔靖.关系数据库的理论及应用.郑州:河南科技出版社,1989.

附录

例 3:设 $U=ABCDE$, $F=\{AB\rightarrow C, AD\rightarrow B, C\rightarrow D, D\rightarrow C\}$,它对应的 FD 图如图 2(a)所示.

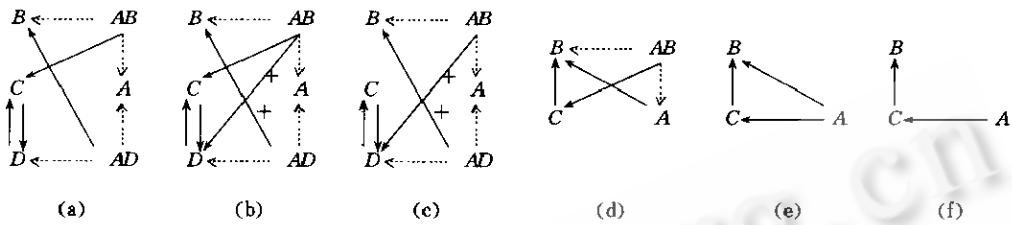

图2

读者可自行验证,图 2(a)的本身就是无冗余覆盖,且结点 AB 和 AD,C 和 D 等价,在合并等价结点后,若直接输出数据库模式为$\{ABCD,CD\}$,但在模式 $ABCD$ 中,非主属性 C 依赖于非候选键 D,故它不是 3NF 的.若按算法 3,增加标识弧后如图 2(b)所示,再消去冗余实弧,冗余结点后如图 2(c)所示,再输出的数据库模式$\{ABD,CD\}$是 3NF 的.

例 4:设 $U=ABC$, $F=\{AB\rightarrow C, A\rightarrow B, C\rightarrow B\}$,它对应的 FD 图如图 2(d)所示.

在图 2(d)中,若先删去冗余实弧,再删去冗余虚弧得图 2(e);若先删去冗余虚弧,再删去冗余实弧则得图 2(f).

EXTERNAL CLOSURE AND ITS APPLICATION IN RELATIONAL DATABASE

Wang Yujun Shi Baile Hu Meichen

(*Department of Computer Science Fudan University Shanghai* 200433)

Abstract In this paper, the authors start with the introduction of generalized FD path and external closure. With the application of them, they show that the number of closures are reduced when calculating the nonredundant cover of a given functional dependency set. Then a new 3NF synthesis algorithm is presented. In this algorithm, the calculation of nonredundant cover in 3NF synthesis is reduced from two passes to one. So the total number of closures that to be calculated are significantly reduced.

Key words Generalized FD path, external closure, nonredundant cover, 3NF synthesis algorithm.