第5卷第3期
1994年3月
软件学报
JOURNAL OF SOFTWARE
Vol. 5, No. 3
Mar. 1994

关于传递闭包和函数依赖完备公理系统存在问题研究*

聂培尧

(山东财政学院,济南 250014)

摘要 数据依赖在数据库设计中起着十分重要的作用.自 Codd 提出函数依赖(FDs)、Fagin 引入多值依赖(MVDs)后,近几年来人们又根据设计中的需要引入多种新的依赖,如在工程数据库设计中所引进的传递闭包依赖(CDs)等.对这些依赖一般是按其是否具有完备的公理系统而划分为两大类,因为完备性公理系统往往具有有效的判定算法为先决条件.本文对 CDs 和 FDs 的 k 元完备公理系统存在问题进行了研究,证明了 CDs 和 FDs 不具有共同的 k 元完备公理系统这一结论.

关键词 数据库设计,函数依赖,传递闭包,公理系统.

近年来,人们对 CAD/CAM 及 CIM 数据库设计中所出现的传递闭包的建模和计算越来越感兴趣[1-4].在数据库的递归查询处理、关系及视图的递归定义问题上传递闭包也起了十分重要的作用[5,6].在传递闭包问题的研究上,对一给定关系上的传递闭包的有效计算方面已做了很多工作并也取得了很大进展[7,8].然而,在对传递闭包的形式化特性及其与函数依赖(FDs)间的相互联系等问题的研究上所做的工作还较少.

最近,G. Gottlob 等人对传递闭包依赖(CDs)的性质进行了较深入的研究[4].在他们的研究中主要探讨了某些闭包依赖是否能根据一给定的 CDs 集合逻辑地推导出来,证明了这一蕴含问题在多项式时间内是可判定的.进一步,他们证明了存在一个能够导出给定 CDs 集合上所蕴含的全部闭包依赖的简单完备公理系统.他们还对 CDs 和 FDs 间的相互联系进行了研究,指出了仅 FDs 不能导出 CDs;CDs 与 FDs 合在一起仅能导出新的 CDs,但不能导出新的 FDs.这些研究结果对数据库中传递闭包特性的理解具有十分重要的意义.本文在此基础上对 FDs 和 CDs 合在一起是否具有一个共同的 k 元完备公理系统问题进行了研究并得出一个否定的结论.

* 本文1991年8月26日收到,1991年11月15日定稿
作者聂培尧,37岁,副教授,主要研究领域为数据库系统,决策支持系统.
本文通讯联系人:聂培尧,济南 250014,山东财政学院

1 基本定义和术语

设一关系模式 R 是全部属性集合 U 的一个子集$\{A_1,A_2,\cdots,A_n\}$且全部的属性具有相同的无限域 O:即对每一 $A\in U$,$\mathrm{Dom}(A)=O$. R 上的一关系 r,记 r(R),是一(可以是无限的)函数 t 的集合:$t:R\to O$. 我们记 $t(A_i)$为 $t[A_i]$并称这些函数为元组.

定义 1.1. 一个关系模式 $R(A_1,A_2,\cdots,A_n)$上的函数依赖(FD)为形如 $X\to Y$ 的表达式,其中 X 和 Y 是 R 的属性集$\{A_1,A_2,\cdots,A_n\}$的子集. 一个 R 上的关系 r 满足 FD $X\to Y$,当且仅当对任意元组 $t,t'\in r$,如果 $t[X]=t'[X]$,必有 $t[Y]=t'[Y]$成立.

定义 1.2. 一个关系模式 R 上的传递闭包依赖,或称为闭包依赖(CD),是把 R 的子集划分成为属性的有序偶对. 因此,一个 CD 可记为$\{(A_1,B_1),(A_2,B_2),\cdots,(A_n,B_n)\}$;其中全部的属性是互不相同的. 一 CD 也可记为 X@Y 的形式,其中 $X=A_1A_2\cdots A_n$,$Y=B_1B_2\cdots B_n$ 是属性序列. 我们称前一种记法为集合记法,后一种记法为顺序记法.

一个关系 r 满足 CD X@Y,其中 $X=A_1A_2\cdots A_n$,$Y=B_1B_2\cdots B_n$,当且仅当对于任意的 t 及 t',若 $t[B_i]=t'[A_i]$,则必存在一元组 t'',使得 $t''[A_i]=t[A_i]$及 $t''[B_i]=t'[B_i]$成立,其中 $i=1,2,\cdots,n$.

注意:在顺序记法中的两个 CDs:$\tau=X@Y$ 及 $\sigma=V@W$ 若在集合记法中表示同一 CD,当且仅当存在一置换 ρ,使得 $\rho(X)=V$ 及 $\rho(Y)=W$. 以上我们假设了所有属性均是定义在相同域上的. 在实际例子中,对于任意的 CD δ 及任意偶对$(A,B)\in\delta$,A 和 B 当然也应当具有相容的域.

令 Σ 为 R 上依赖的集合(FDs 或 CDs),令 σ 为 R 上的一个依赖. 我们说 R 上的任一关系 r 满足 Σ,当且仅当 r 同时满足所有依赖 $\tau\in\Sigma$. 进一步,若对任意(可以是无限)满足 Σ 的关系 r 同时也满足 σ,则称 Σ 逻辑蕴含 σ,记为 $\Sigma\models_i\sigma$(无限蕴含). 若满足 Σ 的每一有限关系 r 也同时满足 σ,则记为 $\Sigma\models_f\sigma$(有限蕴含). 为简单起见,当 i 或 f 已明确时,可简记$\models_i$(或$\models_f$)为$\models$;$\not\models_f$(或$\not\models_i$)为$\not\models$.

我们记 $\Sigma\vdash_{\mathscr{R}}\sigma$ 或 $\Sigma\vdash\sigma$,如果 σ 是使用一推导规则集合 $\mathscr{R}$ 从 Σ 中推导出的.

令 R 为一关系模式,$\mathscr{S}$ 为一依赖集,$\mathscr{S}$ 上的(推导)规则为形如“如果 T 成立,则 τ(if T then τ)”的表达式,其中 T 为 $\mathscr{S}$ 中依赖的有穷集合(每一依赖均称为规则的前件);τ 为 $\mathscr{S}$ 中一个依赖(称为规则的结论). 若 T 中恰好含有 k 个不同的成员,即$|T|=k$,则称该规则为 k 元规则;若 $T=\varnothing$则称 0 元规则. 我们说规则“if T then τ”是有效的,若 $T\models\tau$,即如果 R 上的任一关系满足 T 则必定也满足 τ. 进一步,我们称一(推导)规则的集合 $\mathscr{R}$ 是有效的,如果 $\mathscr{R}$ 中每个成员均为有效的.

令 $\mathscr{R}$ 为 $\mathscr{S}$ 上规则的集合,$\Sigma\subseteq\mathscr{S}$ 为一依赖集,σ 为 Σ 中的一个依赖. 运用 $\mathscr{R}$ 由 Σ 到 σ 的一个推导为 $\mathscr{S}$ 中依赖的有穷序列$\langle\tau_1,\tau_2,\cdots,\tau_m\rangle$;其中 $\tau_m=\sigma$ 且对每一 $i(1\leqslant i\leqslant m)$,或者① $\tau_i\in\Sigma$;或者② 存在 $T\subseteq\{\tau_1,\tau_2,\cdots,\tau_{i-1}\}$使得“if T then τ”为 $\mathscr{R}$ 中的一个规则. 如果存在运用 $\mathscr{R}$ 从 Σ 到 σ 的推导,则记 $\Sigma\vdash_{\mathscr{R}}\sigma$ 或 $\Sigma\vdash\sigma$.

进一步,设 $\mathscr{R}$ 为 $\mathscr{S}$ 上的一规则集,若对任意 $\Sigma\subseteq\mathscr{S}$ 和 $\sigma\in\mathscr{S}$ 有:若 $\Sigma\models\sigma$ 则 $\Sigma\vdash_{\mathscr{R}}\sigma$,则称公理系统 $\mathscr{R}$ 是完备的. 这里的定义是文献[9,10]正确性和完备性的综合. 如果规则集 $\mathscr{R}$

中每一规则 $\rho\in\mathscr{R}$ 至多是 k 元的，即 ρ 为 r 元规则，$r\leqslant k$，则称 $\mathscr{R}$ 为 k 元公理系统.

在以下的讨论中，我们使用符号≡表示属性值的句法等价；符号＝表示属性值的等价.

2　FDs 与 CDs 完备公理存在问题

为了讨论 FDs 与 CDs 完备公理存在问题，我们先给出文献[4]中的两个结论，在此我们以定理的形式给出.

定理 2.1. FDs 自身不蕴含 CDs.

证明：见文献[4]中定理 4.5. □

定理 2.2. 令 D 为 FDs 和 CDs 的集合，除了那些由 FDs 集合中导出的 FDs 外，不再有任何其它的 FDs 可由 D 中逻辑地导出.

证明：见文献[4]中定理 4.6. □

以上定理说明了 FDs 和 CDs 合在一起不能导出新的 FDs.

下面我们来讨论 FDs 和 CDs 的 k 元完备公理系统的存在问题.

令 R 为一关系模式，$\mathscr{S}$ 为一依赖集合. 若 $\tau=X@Y$，我们记 τ' 为 $Y@X$，记 $\tau\equiv\sigma$ 当且仅当 $\tau\models\sigma$ 及 $\sigma\models\tau$. 若 τ 和 σ 为 CDs，则由 CDs 的完备性结果[4]可以证明，$\tau\equiv\sigma$ 当且仅当 $\tau=X@Y$ 及 $\sigma=Y@X$.

定义 2.1. $\tau\in\mathscr{S}$ 称为与 $\sigma\in\mathscr{S}$ 无关的，当且仅当对任意不包含任何 $\sigma'=\sigma$ 的 $\Delta(\Delta\subseteq\mathscr{S})$ 有：如果 $\Delta\models\sigma$，则 $\Delta-\{\tau\}=\sigma$；否则，τ 称与 σ 相关.

定义 2.2. CD X@Y 的度称为 X，Y 的长度.

引理 2.1. 令 $\tau\in\mathscr{S}$，$\tau\not\equiv\sigma$. τ 与 σ 无关当且仅当对任意 $\Delta\subseteq\mathscr{S}$，$\Delta\not\models\sigma$ 蕴含 $\Delta\cup\{\tau\}\not\models\sigma$.

证明：直接可得. □

我们定义算子 Set 如下，该算子把有序偶对的集合映象到一无序偶对集合，即映象到一个具有两个元素的集合的集合上.

$\mathrm{Set}:\{(x_1,y_1),\cdots,(x_n,y_n)\}\rightarrow\{\{x_1,y_1\},\cdots,\{x_n,y_n\}\}$.

令 τ 为形如 X@Y 的 CD，τ^c 表示 CD 的集合：$\{\delta|\mathrm{Set}(\delta)\subseteq\mathrm{Set}(\tau)\}$.

定理 2.3. 令 $\sigma=X@Y$，则每一 CD $V@W\in\sigma^c$ 都是与 σ 无关的.

证明：证明的思路为，使用引理 2.1，我们证明对于任意的 $\tau\notin\sigma^c$，若 $\Delta\subseteq\mathscr{S}$ 不蕴含 σ，则 $\Delta\cup\{\tau\}$ 也不蕴含 σ.

令 Δ，$\sigma=X_1\cdots X_n@Y_1\cdots Y_n$，$\tau=V@W$. 若 $\Delta\not\models\sigma$，则存在一 $R=X_1\cdots X_nY_1\cdots Y_nZ_1\cdots Z_m$ 上的关系 r，使得 r 满足 Δ 但不满足 σ. 进一步，根据关系 r 我们构造一关系 r′并证明① r′满足 Δ；② r′满足 τ；③ r′不满足 σ，从而得到 $\Delta\cup\{\tau\}\not\models\sigma$. 我们说关系 r′与 r 是相同的，除了 X_i 列及 Y_i 中所有标有 $i(i=1,2,\cdots,n)$ 的项和某些标有 $j(j=1,2,\cdots,m)$ 以及 1 的 Z_j 项外. 也即，r′是通过以下映象 h 把 r 元组映射到 r′元组而得到的：

$h:t\in r\rightarrow t'\in r'$ 使得 $t'[X_i]=t[X_i]_i$，$t'[Y_i]=t[Y_i]_i$，$(i=1,2,\cdots,n)$ 以及任 $t'[Z_j]=t[Z_j]'$，$(j=1,2,\cdots,m)$.

下面我们证明，若 r 满足某些 $\sigma\in\Delta$，则 r′满足 σ. 由 r 的定义可以推出，对任意的 $A\in R$，$u[A]=v[A]$ 当且仅当 $u'[A]=v'[A]$，其中 $u'=h(u)$，$v'=h(v)$. 由此可见，映射 h 在不同元

组的相同列中保持了值的相等，所以它保持了 FDs 与 CDs 的有效性.

我们下面证明 r'满足 $\tau=V@W$. 由映射 h 的定义，$r'[V_k]=r'[W_k]$，$(k=1,2,\cdots,n)$当且仅当存在某些 $i\in\{1,2,\cdots,n\}$使得$(V_k=X_i$ 且 $W_k=Y_i)$或者$(W_k=X_i$ 且 $V_k=Y_i)$成立. 根据假设 $\tau\notin\sigma^c$ 和 $Set(\tau)\not\subseteq Set(\delta)$，所以不存在两个不同的元组 $u',v'\in r'$使得 $u'[W]=v'[V]$.

我们再证明 r'不满足 σ. 由初始假设 r 不满足 σ，则存在 $u,v\in r$ 使得 $u[Y]=v[X]$但不存在一 $Z\in r$ 使得 $Z[X]=u[X]$及 $Z[Y]=v[Y]$. 考虑 $u'=h(u)$及 $v''=h(v)$，由 h 的定义，$u'[Y]=v'[X]$. 进一步，对于任意 $Z'\in r$，$Z'[X]=u'[X]$及 $Z'[Y]=v'[Y]$不成立. 因为映象 h 在每一列中均可保持其值的特性. □

定理 2.4. 令 Δ 为 CDs 与 FDs 的集合，$\Delta\models X@Y$，$X@Y\in\Delta$ 且 $Y@X\notin\Delta$，则在 Δ 中存在一 CD V@W，使得 X@Y 的度大于或等于 V@W 的度.

证明：因为仅 FDs 不能蕴含 CDs（定理 2.1）且在 Δ 中存在某些与 X@Y 相关的 CD V@W. 再根据定理 2.2 中的结论立即可得. □

关于 FDs 及 CDs 各自完备公理系统存在性问题，文献[4]及[11]已分别证明了 CDs 和 FDs 分别具有各自完备的公理系统. 下面我们来考虑 CDs 和 FDs 合在一起的完备公理系统存在问题. 本文以下的讨论对这一问题给出了一个否定的结论.

令 Σ^+ 为集合$\{\delta\in\mathscr{S}\mid\Sigma\models\sigma\}$，我们有以下定理.

定理 2.5. 令 $\Sigma\subseteq\mathscr{S}$，$\sigma\in\mathscr{S}$，$\Sigma\models\sigma$，并且对所有的 $\tau\in\Sigma$：$\tau\neq\sigma$. 若 Σ 具有一 k 元完备公理系统 $\mathscr{R}$，则存在一 $\Delta\subseteq\mathscr{S}$ 使得 $\Delta\models\sigma$，其中$|\Delta|\leqslant k$，$\Sigma\models\Delta$，并且对所有的 $r\in\Delta$：$\tau\neq\sigma$ 成立.

证明：$\Sigma\models\sigma$，$\mathscr{R}$ 是完备的，因而有 $\Sigma\vdash_{\mathscr{R}}\sigma$，并存在一从 Σ 到 σ 的推导$\langle\tau_1,\tau_2,\cdots,\tau_m=\sigma\rangle$；其中对于 $i\neq m$ 有 $\tau\neq\sigma$. 考虑推导过程的第 m 步，存在一集合 $T\subseteq\{\tau_1,\tau_2,\cdots,\tau_{m-1}\}$使得“if T then σ”为 $\mathscr{R}$ 中的一规则. 显然，$\Sigma\models T$，因为 $\mathscr{R}$ 中的每一规则均为 k 元的，$|T|\leqslant k$，$T\equiv\Delta$. □

令 $\Sigma_\sigma^+=\{\tau\in\Sigma^+\mid\tau$ 与 σ 相关 $\}$，我们有以下推论：

推论 2.6. 对于定理 2.5 中的 Δ，Σ 及 σ，存在一 $\Delta'\subseteq\Delta$ 使得 $\Delta'\subseteq\Sigma_\sigma^+$ 及 $\Delta'\models\sigma$.

证明：$\Delta'=\Delta-\{\tau\in\Delta\mid\tau$ 与 σ 无关$\}$ □

定理 2.7. 不存在一非负整数 k，使得对 CDs 和 FDs 存在共同的 k 元完备公理系统.

证明：设 k 为一固定的自然数，令 R 为一具有 $2(k+1)$个不同属性的关系模式，即：

$R=A_1\cdots A_{k+1}B_1\cdots B_{k+1}$；令 Σ 为集合：

CD_1 $A_1A_3A_4\cdots A_kA_{k+1}@B_1B_3B_4\cdots B_kB_{k+1}$

FD_1 $A_1A_3A_4\cdots A_kA_{k+1}\rightarrow B_1$

CD_2 $A_1A_2A_4\cdots A_kA_{k+1}@B_1B_2B_4\cdots B_kB_{k+1}$

FD_2 $A_1A_2A_4\cdots A_kA_{k+1}\rightarrow B_2$

…

CD_i $A_1\cdots A_iA_{i+2}\cdots A_{k+1}@B_1\cdots B_iB_{i+2}\cdots B_{k+1}$

FD_i $A_1\cdots A_iA_{i+2}\cdots A_{k+1}\rightarrow B_i$

…

CD_{k-1} $A_1A_2A_3\cdots A_{k-1}A_{k+1}@B_1B_2B_3\cdots B_{k-1}B_{k+1}$

FD_{k-1} $A_1A_2A_3\cdots A_{k-1}A_{k+1}\rightarrow B_{k-1}$

CD_k　$A_1A_2A_3\cdots A_{k-1}A_k@B_1B_2B_3\cdots B_{k-1}B_k$

FD_k　$A_1A_2A_3\cdots A_{k-1}A_k\to B_k$

CD_{k+1}　$A_2A_3A_4\cdots A_kA_{k+1}@B_2B_3B_4\cdots B_kB_{k+1}$

FD_{k+1}　$A_2A_3A_4\cdots A_kA_{k+1}\to B_{k+1}$

令 σ 为 CD $X@Y$，$X=A_1A_2\cdots A_{k+1}$，$Y=B_1B_2\cdots B_{k+1}$，令为 $\mathscr{S}'_F$ 为 R 上所有侯选 FD 的集合，并且令 $\mathscr{S}_c$ 为 R 上所有侯选 CD 的集合. 我们有以下的断言：

断言 1. $\Sigma\models\sigma$.

令 CD_i^s 为 CDs $\{\tau\,|\,Set(\tau)=Set(CD_i)\}$ 的集合，对于 $i\neq j$ 有 $CD_i^s\cap CD_j^s=\varnothing$. 进一步，令 Σ_σ^+ 为集合 $\{\tau\in\Sigma^+\,|\,\tau$ 与 σ 相关$\}$. 再令 Σ_F 为 Σ 中 FDs 的集合，Σ_c 为 Σ 中 CDs 的集合且 Σ_c^s 为 CDs $\{\tau\in\mathscr{S}\,|$ 存在一 $\sigma\in\Sigma_c$ 使得 $Set(\tau)=Set(\delta)\}\cup\{\tau\,|\,\tau\not\equiv\sigma, Set(\tau)=Set(\sigma)\}$ 的集合. 我们有断言 2：

断言 2. 不存在一 $\Delta\subseteq\Sigma_\sigma^{+'}$ 使得 $|\Delta|\leqslant k$ 且 $\Delta\models\sigma$. 由于篇幅所限，断言 1 和断言 2 的证明从略.

断言 3. $\Sigma_F^+\cup\Sigma_c^s=\Sigma_\sigma^{+'}\cup\Sigma_\sigma^+$.

下面我们给出断言 3 的证明.

证明：我们要证明，如果 $\Sigma\models\tau$ 并且 τ 与 σ 相关，则 $\tau\in\Sigma_\sigma^{+'}$.

假设 τ 是一 FD，由定理 2.2 可知，$\Sigma^+\cap\mathscr{S}_F=\Sigma_F^+$. 由 $\Sigma\models\tau$ 可得 $\tau\in\Sigma^+$，$\tau\in\Sigma_F^+$ 及 $\tau\in\Sigma_\sigma^+$. 假设 τ 是一 CD，Σ 中的所有 CDs 均为 k 度，由定理 2.3，若 $Set(V@W)\subseteq Set(\delta)$，则 CD $V@W$ 仅与 σ 相关. 因此，τ 的度为 k 或 k+1. 根据定义，所有相关的度为 k 及 k+1 的 CDs 均属于 Σ_c^s，因此，$\tau\in\Sigma_\sigma^{+'}$. 这就证明了断言 3. □

下面我们来完成定理 2.7 的证明. 由断言 1 可知 $\Sigma\models\sigma$，再由断言 2 和断言 3，对于每一 $\Delta\subseteq\Sigma_\sigma^+\subseteq\Sigma_\sigma^{+'}$，$|\Delta|\leqslant k$，$\Delta\models\sigma$ 不成立. 因此，由推论 2.6 可得，对于 FDs 和 CDs，不存在一共同的 k 元完备公理系统. □

结束语：本文对 FDs 及 CDs 是否具有共同的 k 元完备公理系统进行了研究并得出了否定的结论，这一结果对于了解数据库中传递闭包依赖的特性是有意义的. 传递闭包依赖是为工程数据库设计所引进的一种新的数据依赖，而其完备的公理系统往往对设计有效的依赖判定算法起着重要的作用. 一般来说，一种不具备完备公理系统的依赖，其许多判定问题是困难的（非多项式的），甚至根本找不到确定的判定过程. 对这种依赖的判定问题，则应力求通过找出有效的近似算法来处理.

参考文献

1 Ioannidis Y E. On the computation of the transitive closure of relational operators. Proc. of the 12th Inter. Conf. on VLDB, Kyoto, 1986.

2 H Lu. New strategies for computing the transitive closure of database relation. Proc. of the 13th Inter. Conf. on VLDB, Brighton, 1987.

3 H Lu. Mikkilieni K, Richardson J P. Design and evaluation of algorithms to compute the transitive closure of a database relation. Proc. IEEE 3rd Inter. Conf. on Data Engineering, Los Angeles, 1987.

4 Gottlob G *et al*. On the interative between transitive closure and functional dependencies. MFDB'89, Springer-Verlag, 1989.

5 Agrawal R. Alpha: an extension of relational algebra to express a class of recursive queries. Proc. IEEE 3rd Inter. Conf. on Data Engineering, Los Angeles, 1987.

6 Jagadish H V. Agrawal R. A study of transitive closure as a recursion mechanism. Proc. of ACM SIGMOD, San Francisco, 1987.

7 Warshall S. A theorem on boolean matrices. J. ACM, 1962; **9**(1).

8 Warshall S. A modification of warshall's algorithm for the transitive closure of binary relations. Comm. ACM, 1975; **18**(4).

9 Ullman J D. Principles of database systems. 2nd Edition, Computer Science Press Inc., 1982.

10 Maier D. The theory of relational database. Computer Science Press, Rockville Md., 1983.

11 Casanova M A, Fagin R, Papdimitrion C H. Inclusion dependencies and their interaction with functional dependencies. J. Computer System Science, 1984; **28**(1).

ON THE EXISTENCE OF FINITE COMPLETE AXIOMATIZATION FOR TRANSITIVE CLOSURE AND FUNCTIONAL DEPENDENCIES

Nie Peiyao

(*Shandong Financial Institute, Jinan* 250014)

Abstract Closure dependencies, or CDs, are introduced to capture formally transitive closure relationships between attributes of a relational schema. The problem of a finite complete axiomatization for the interaction of CDs with functional dependencies (FDs) is investigated, leading to a negative result for the existence of a k-ary axiomatization for CDs and FDs taken together, in which every rule is k-ary for some fixed k.

Key words Database design, functional dependencies, transitive dosure, axiomatization systems.