数理解析研究所講究録
421巻 1981年 106-120

106

多値従属を考慮した関係表現における制約の導出

大阪大学基礎工学部　伊藤 実　岩崎 元昭
谷口 健一　嵩 忠雄

1. まえがき

関係データベース（relational database）における質問（query）を表現する言語、すなわち、関係言語の一つとして関係代数（relational algebra）が知られている。Rを、s個の関係スキーム（relational scheme）$R_1, \cdots, R_s$ からなる関係データベーススキーマ（relational database schema）とする。関係代数による質問は、オペランドに関係スキーム名をもち、オペレータとしてその関係代数で定義されている基本演算をもつ一つの式（以後、関係表現（relational expression）という）で表される。基本演算として、関係和（union）、関係差（set difference）、直積（cross product）、射影（projection）、及び、選択（selection）を考える。（関係積（intersection）や結合（join）は、それらの基本演算の組合せで実現できる[9]。）本論文では、選択演算として、(1)特定の成分（属性）に指定

された定数をもつ組だけを関係から選ぶものと、(2)特定の2つの成分が同じ値をもつ組だけを関係から選ぶものを考える。

R中の各関係スキームRiは一般にその関係例(relation instance)が満すべき制約(従属性(data dependency)等)が指定されている。Rのデータベース$I = \{r_1, \cdots, r_s\}$に対して、特に各r_i($1 \leq i \leq s$)がR_iの関係例ならば、IをRのデータベース例(database instance)という。R上の関係表現Eは、Rの各データベースIからそれに対応する一つの関係E(I)への写像を定義する。Eを利用者による質問とすると、E(I)はその質問に対する利用者ビュー(user's view)と考えられる。本論文では、R(すなわち、各R_iの属性数と制約集合)、及び、R上の関係表現Eと、一つの制約dが任意に与えられたとき、その制約dがEで成立するかどうか、すなわち、R上の任意のデータベース例Iに対して関係E(I)がその制約dを常に満たすかどうかの判定問題について考察する。Eで成立する制約を求めることは、質問処理、特に、ビュー上での利用者によるデータの更新処理を行うのに重要である。例えば、E(I)に対して、利用者がEで成立する制約に反する更新を行おうとすれば、システムはデータベースの統合性(integrity)を保つためにそのような更新を拒否しなければならないことがある。

上記の問題に対して、文献(5)に次の結果が示されている。「$\mathbb{R}$に属する各関係スキームR_iに対して、関数従属(functional dependency;FD)[2]の集合F_iが指定されるとする。このとき、

(イ) Eを$\mathbb{R}$上の任意の関係表現とし、dを任意のFDとすると、dがEで成立するかどうかの判定問題は決定不能である。

(ロ) Eを関係差を含まない$\mathbb{R}$上の任意の関係表現とし、dを任意のFD、又は、value equality (VEQ)、又は、domain equality (DEQ)とすると、dがEで成立するかどうかの判定問題は決定可能である。但し、VEQとは、$A \equiv c$なる文で、関係に属する各組のA成分が常に定数cをもつという制約を表す。また、DEQとは、$A = A'$なる文で、関係に属する各組のA成分とA'成分の値が常に等しいという制約を表す。」

本論文では、上記(ロ)の結果を多値従属(multivalued dependency;MVD)[2]に拡張したときの結果を示す。

データベースを設計する手法として、関係スキームRを幾つかのサブスキームの集合$\{R_1, \cdots, R_s\}$に分解する場合、その分解によって元のRで成立しなければならない制約が保存される必要がある。これは制約保存条件(constraint preservation condition)[7]と呼ばれ、「$\{R_1, \cdots, R_s\}$の任意の関係例の集合を$\{r_1, \cdots, r_s\}$とすると、$r_1 \bowtie \cdots \bowtie r_s$はRの関係例であるか。」と

いう問題として表すことができる。但し、⋈は自然結合(natural join)[9]を表す。これは、Rで指定される各制約が、$R_1 \bowtie \cdots \bowtie R_s$で成立するかどうかの判定問題に帰着できる。R及び$R_1, \cdots, R_s$で指定される制約がFD及びMVDの場合、この判定問題の決定可能性は、文献[7]に未解決の問題の一つとして挙げられていたが、本論文の結果を用いれば、任意に与えられたFDスはMVDが$R_1 \bowtie \cdots \bowtie R_s$で成立するかどうかの判定問題が多項式時間で解けることが示せる。

2. 諸定義

値の可算無限集合を$\Gamma = \{c_1, c_2, \cdots\}$とする。次数(degree)mの関係(relation)rとは、Γに属する値を成分としてもつようなm字組(m-tuple)の有限集合である。本論文では、関係の直積や和集合をとる演算を考えるので、関係に属する組の各成分を、通常の属性(attribute)を用いて識別するよりも、成分番号(domain number)を用いて何番目の成分であるかという形式で識別する方が便利である。

$U = \{1, \cdots, m\}$とおく。関係rに属する組t、及び、Uに属する番号Aに対して、tのA成分をt[A]と表す。更に、Uの部分集合$X = \{A_1, \cdots, A_k\}$に対して、$\langle t[A_1], \cdots, t[A_k] \rangle$をt[X]と表す。特に各成分の順番が問題になる場合(例え

ば，組の成分の置換）には，$\sigma(X) = A_{i_1} \cdots A_{i_k}$ なる系列化関数 σ を定義し，組 $\langle t[A_{i_1}], \cdots, t[A_{i_k}] \rangle$ を $t[\sigma(X)]$ と表す。但し，$(i_1, \cdots, i_k)$ は $(1, \cdots, k)$ のある置換（permutation）を表す。以下では，番号の集合 X，Y に対して，その和集合 $X \cup Y$ を XY と略記することがある。次に，本論文で使用する関係に対する5種類の演算を定義する。

r を U 上の関係とし，X を U の部分集合とする。このとき，関係 r の $\sigma(X)$ 上への射影（projection）とは，$\{t[\sigma(X)] \mid t \in r\}$ なる関係で，$r[\sigma(X)]$ と表す。但し，σ は X に対する系列化関数である。σ を U に対する系列化関数とすれば，射影 $r[\sigma(U)]$ によって r の各列の置換が行える。

r_1 を次数 m の関係とし，r_2 を次数 n の関係とする。このとき，r_1 と r_2 の直積とは，$\{t_1 \cdot t_2 \mid t_1 \in r_1$，且つ，$t_2 \in r_2\}$ なる次数 $m+n$ の関係であり，$r_1 \times r_2$ と表す。但し，$t_1 \cdot t_2$ は t_1 と t_2 の連接，すなわち，$t_1 = \langle a_1, \cdots, a_m \rangle$，$t_2 = \langle b_1, \cdots, b_n \rangle$ とおけば，$t_1 \cdot t_2 = \langle a_1, \cdots, a_m, b_1, \cdots, b_n \rangle$ である。同様に，次数 m_1 の関係 r_1，…，次数 m_k の関係 r_k（$r_1, \cdots, r_k$ はすべて異なる関係である必要はない）に対して，$r_1 \times \cdots \times r_k$ は $\{t_1 \cdots t_k \mid t_1 \in r_1, \cdots, t_k \in r_k\}$ なる次数 $\sum_{i=1}^{k} m_i$ の関係である。このとき，r_i に属する組 t_i の j 番目の成分は，$r_1 \times \cdots \times r_k$ では $(\sum_{l=1}^{i-1} m_l + j)$ 番目の成分であるが，簡単のため，$\sum_{l=1}^{i-1} m_l + j$ を

$j^{(i)}$ と略記する。同様に、$U_i = \{1, \cdots, m_i\}$ とすると、U_i の任意の部分集合 $X = \{A_1, \cdots, A_m\}$ に対して、$X^{(i)} = \{A_1^{(i)}, \cdots, A_m^{(i)}\}$ とする。

r_1, r_2 をそれぞれ次数 m の関係とする。このとき、r_1 と r_2 の次数が等しいので、r_1 と r_2 の関係和が定義できる。r_1 と r_2 の関係和とは、$\{t \mid t \in r_1$, 又は, $t \in r_2\}$ なる次数 m の関係で $r_1 \cup r_2$ と表す。もし、r_1 と r_2 の次数が異なれば、$r_1 \cup r_2$ は定義されない。k 個の関係 $r_1, \cdots, r_k$ がすべて同じ次数をもつならば、それらの関係和を $r_1 \cup \cdots \cup r_k$ と表す。

選択には2種類の演算 value equality (VEQ) と domain equality (DEQ) がある。r を次数 m の関係とし、$U = \{1, \cdots, m\}$ とする。VEQとは $A \equiv c$ なる文である。但し、$A \in U$, 且つ、$c \in \Gamma$ である。このとき、$r[A \equiv C]$ を $\{t \mid t \in r$ 且つ $t[A] = c\}$ なる関係とする。VEQの集合 $\{A_1 \equiv c_1', \cdots, A_n \equiv c_n'\}$ をまとめて、$A_1 \cdots A_n \equiv c_1' \cdots c_n'$ と書くことがある。DEQとは $A = A'$ なる文である。但し、$A, A' \in U$ である。このとき、$r[A = A']$ を $\{t \mid t \in r$ 且つ $t[A] = t[A']\}$ なる関係とする。DEQの集合 $\{A_{i_1} = A_{j_1}, \cdots, A_{i_n} = A_{j_n}\}$ をまとめて、$A_{i_1} \cdots A_{i_n} = A_{j_1} \cdots A_{j_n}$ と書くことがある。

関数従属 (FD) とは、$X \to A$ なる文である。但し、$X \subseteq U$ 且つ $A \in U$ である。次数 m の関係 r に属する任意の組 t_1, t_2

に対して、もし $t_1[X] = t_2[X]$ ならば、$t_1[A] = t_2[A]$ であるとき、関係 r は $X \to A$ を満たすという。FDの集合 $\{X \to A_1, \cdots, X \to A_m\}$ をまとめて $X \to A_1 \cdots A_m$ と表すことがある。

多値従属(MVD)とは、$X \twoheadrightarrow Y$ なる文である。但し、$X, Y \subseteq U$ である。$Z = U - XY$ とおく。次数 m の関係 r に属する任意の組 t_1, t_2 に対して、もし $t_1[X] = t_2[X]$ ならば、$t_3[XY] = t_1[XY]$ 且つ $t_3[XZ] = t_2[XZ]$ なる組 t_3 も r に属するとき、関係 r は $X \twoheadrightarrow Y$ を満たすという。

ある組 t に対して、$t[A] = c$ であるとき、t は VEQ $A \equiv c$ を満たすという。更に、関係 r に属する任意の組 t が $A \equiv c$ を満たすとき、関係 r は $A \equiv c$ を満たすという。同様に、ある組 t に対して、$t[A] = t[A']$ であるとき、t は DEQ $A = A'$ を満たすといい、関係 r に属する任意の組 t が $A = A'$ を満たすとき、関係 r は $A = A'$ を満たすという。

r をある関係とし、M をMVDの集合とする。このとき、r に対する M のもとでの chase とは、r(に属する組)を含み、且つ、M を属するすべてのMVDを満たす最小の関係で、r^* と表す。r^* を実際に求める手続きは文献(6)に書かれている。r に含まれる組の個数を k とすれば、一般に r^* を記述するのに $O(k^{deg(r)})$ の領域が必要である。但し、$deg(r)$ は関係 r の次数を表す。

関係スキームを$R\langle m, D\rangle$と表す。但し、Rはスキーム名、mは次数、DはFDおよびMVDの集合とする。次数mの関係rがDに属するすべてのFDおよびMVDを満たすならば、その関係rをRの関係例（rはRの制約を満たす関係）という。関係スキームの集合$\mathbb{R} = \{R_1\langle m_1, D_1\rangle, \cdots, R_s\langle m_s, D_s\rangle\}$を関係データベーススキーマという。各関係$r_i$が次数$m_i$であるような関係の集合$I = \{r_1, \cdots, r_s\}$を$\mathbb{R}$のデータベースといい、更に、各$r_i$が$R_i$の関係例であれば、I を$\mathbb{R}$のデータベース例（rは$\mathbb{R}$の制約を満たすデータベース）という。

$\mathbb{R}$上の関係表現とは、オペランドに$\mathbb{R}$に属する関係スキーム名をもち、オペレータとして射影、直積、関係和、VEQおよびDEQをもつ一つの式である。従って、R_1, $R_2[\sigma_1(X)]$, $(R_1[\sigma_2(Y)])\times R_2\times R_3$, $R_1 \cup (R_2[\sigma_3(Z)])$, $R_1[A\equiv C]$, $R_2[A=A']$等はすべて$\mathbb{R}$上の関係表現である。Eを$\mathbb{R}$上の関係表現とし、$I = \{r_1, \cdots, r_s\}$を$\mathbb{R}$のデータベースとする。このとき、Eは I から一つの関係E(I)への写像を定義する。その値は、Eのオペランドに現れる各関係スキーム名R_iに関係r_iを代入することによって得られる。以下では、射影、VEQおよびDEQは直積よりも強い演算で、且つ、直積は関係和よりも強い演算とみなして、誤解のない限り関係表現に現れる括弧を省略する。$\mathbb{R}$上の関係表現Eに対して、Eの次数（deg(E)と表

す）を、関係E(I)の次数と定義する。

Rに属する関係スキーム$R_i\langle m_i, D_i\rangle$及び一つの制約（すなわち、FD，MVD，VEQ又はDEQ）dに対して、もしR_iの任意の関係例がdを常に満たすならば、dはR_iで成立する（valid）という。同様に、R上の関係表現Eが与えられたとき、もしRの制約を満たすデータベースIに対して、E(I)が制約dを常に満たすならば、dはEで成立するという。

Eを関係和を含まないR上の関係表現とする。このとき、Eは$(R_{k_1}\times\cdots\times R_{k_m})[\xi\equiv V][P=Q][\sigma(W)]$という標準形に変換できることが知られている。[5] この変換は$O(|E|^2+deg(E))$時間で容易に行える。但し、|E|はEの記述の長さを表す。Eの標準形から射影を除いた$(R_{k_1}\times\cdots\times R_{k_m})[\xi\equiv V][P=Q]$をEのフレーム（frame）という。

3. 結果

〔定理1〕　$R=\{R_1\langle m_1, D_1\rangle, \cdots, R_s\langle m_s, D_s\rangle\}$を関係データベーススキーマとする。但し、各$D_i$は任意のFD及びMVDの集合とする。このとき、R上の任意の関係表現E及び任意の制約dが与えられたとき、dがEで成立するかどうかの判定問題は決定可能である。　□

〔定理2〕 $\mathbb{R} = \{R_1\langle m_1, F_1\rangle, \cdots, R_s\langle m_s, F_s\rangle\}$ を任意の関係データベーススキーマとする。但し、各 F_i はFDの集合とする。Eを $\mathbb{R}$ 上の任意の関係表現とし、dを任意の制約（FD, VEQ又はDEQ）とする。このとき、dがEで成立しないかどうかの判定問題はNP完全である。 □

$\mathbb{R} = \{R_1\langle m_1, D_1\rangle, \cdots, R_s\langle m_s, D_s\rangle\}$ を関係データベーススキーマとする。但し、各 R_i に対して、$D_i = F_i \cup M_i$ とし、$U_i = \{1, \cdots, m\}$ とおく。まず、2,3の定義を与える。Eを関係和を含まない $\mathbb{R}$ 上の任意の関係表現とし、$U = \{1, \cdots, deg(E)\}$ とおく。

Eで成立するすべてのDEQによってUは同値類 $\mathcal{L}(E)$ に分割される。すなわち、Uの分割 $\mathcal{L}(E) = \{L_1, \cdots, L_p\}$ は、「任意のA, $A' \in U$ に対して、$A = A'$ がEで成立するとき、かつそのときに限り、AとA'は $\mathcal{L}(E)$ の同じブロックに属する（すなわち、あるiが存在して、$A \in L_i$ 且つ $A' \in L_i$）」という性質をもつ。VEQ $A \equiv C$ がEで成立するための必要十分条件は、ある $A' \equiv C$ なるVEQがEに現れており、且つ、AとA'が $\mathcal{L}(E)$ の同じブロックに属することである。従って、分割 $\mathcal{L}(E)$ が求まれば、任意のDEQ又はVEQがEで成立するかどうかの判定が行える。$\mathcal{L}(E)$ の同じブロックに属す

るA, A'に対して、A≡C及びA'≡C'がEに現れており、且つ、C≠C'ならば、Eは、Rの任意のデータベースIに対してE(I)=φなる関係表現である。式(E)が求まれば、この判定は可能である。

XをUの部分集合とする。左辺がXで、且つ、Eで成立するFDの右辺全体の集合をXを閉包(closure)といい、𝒞(X, E)と表す。すなわち、𝒞(X, E)={A | X→AはEで成立}と定義する。また、X↠BはEで成立し、且つ、任意のB'(φ≠B'⊊B)に対してX↠B'はEで成立しないような右辺Bの集合をXの従属基(dependency basis)といい、ℳ(X, E)と表す。ℳ(X, E)はUの分割で、X↠YがEで成立するための必要十分条件は、Yがℳ(X, E)に属する幾つかのブロックの和集合で表されることである。従って、𝒞(X, E)(又は、ℳ(X, E))が求まれば、左辺がXの任意のFD(又はMVD)が成立するかどうかの判定が行える。関係スキーム<R, D>に対して、もしDがFD及びMVDの集合ならば、任意のX⊆{1, …, m}に対して𝒞(X, R)及びℳ(X, R)を求めるO(m·‖D‖)時間の手続きが知られている。(8) 但し、‖D‖はDの記述の大きさを表す。

関係和を含まないR上の任意の関係表現は、多項式時間で$(R_{k_1} \times \cdots \times R_{k_m})[Z \equiv V][P = Q][W]$なる標準形に変形できる。

$E=(R_{B_1}\times\cdots\times R_{B_m})[\xi\equiv V][P=Q]$、且つ、$U=\{1,\cdots,deg(E)\}$とおく。

〔定理3〕　次の条件1又は条件2が成立するとき、$\mathcal{L}(E)$が（R，Eの記述長に関して）多項式時間で求まる。すなわち、任意のVEQ又はDEQがEで成立するかどうかの判定は多項式時間で行える。

（条件1）　各M_i（$1\leqq i\leqq s$）が空集合，すなわち，各関係スキームでMVDが制約として与えられていない。

（条件2）　E中に各関係スキーム名が高々2回しか現れない。　□

〔定理4〕　次の条件3が成立するとき，任意の$X\subseteq U$に対して$\mathcal{F}(X,E)$及び$\mathcal{M}(X,E)$が多項式時間で求まる。すなわち、任意のFD又はMVDがEで成立するかどうかの判定は多項式時間で行える。

（条件3）　E中にDEQが現れない。すなわち，$E=(R_{B_1}\times\cdots\times R_{B_m})[\xi\equiv V]$という形をしている。　□

〔定理5〕　$\mathcal{L}(E)$が既知ならば、任意の$X\subseteq U$に対して、$\mathcal{F}(X,E)$及び$\mathcal{M}(X,E)$はR，E，X，$\mathcal{L}(E)$の記述長に関して多項式時間で求まる。従って、定理3の結果より、条件1

又は条件2が成立するとき，任意のFD又はMVDがEで成立するかどうかの判定は多項式時間で行える。 □

〔定理6〕 上記のすべての定理は，Eが射影を含む場合にも容易に拡張できる。この結果の応用として，任意のFD又はMVDが$R_1 \bowtie \cdots \bowtie R_s$で成立するかどうかの判定は多項式時間で行える。 □

6. あとがき

本論文で示した手法を用いて，次のような拡張は容易に行える。

(1) Rに属する各関係スキームR_iの制約の集合D_iがFD及び結合従属(join dependency：JD)[6]からなる集合とする。JDはMVDの一つの一般化であり，特別な場合としてMVDを含む。このとき，R上の任意の関係表現E，及び，任意のFD，JD，VEQ又はDEQ dが与えられたとき，dがEで成立するかどうかの判定問題は決定可能である。

(2) 上記(1)の結果に加えて，関係表現Eに現れる演算として，関係積を許しても，dがEで成立するかどうかの判定問題は決定可能である。

残された問題の一つは、R上の任意の関係和を含まない関係表現Eに対して、同値類 $\mathcal{M}(E)$ を求める能率のよい手続きを（もし存在するならば）見つけることである。

参考文献

(1) Beeri, C. On the membership problem for functional and multivalued dependencies in relational databases. ACM Trans. Database Syst. 5, 3 (Sept. 1980), 241-259.

(2) Beeri, C., Fagin, R. and Howard, J. H. A complete axiomatization for functional and multivalued dependencies in database relations. Proc. 3rd ACM SIGMOD Int. Conf. Management of Data, Toronto, 1977, pp. 47-61.

(3) Hagihara, K., Ito, M., Taniguchi, K. and Kasami, T. Decision problems for multivalued dependencies in relational databases. SIAM J. Compt. 8, 2 (May 1979), 247-264.

(5) Klug, A. Calculating constraints on relational expressions. ACM Tras. Database Syst. 5, 3 (Sept. 1980), 260-290.

(6) Maier, D., Mendelzon, A. O. and Sagiv, Y. Testing implications of data dependencies. ACM Trans. Database Syst. 4,4 (Dec. 1979) 455-469.

(7) Maier, D., Mendelzon, A., Sadri, F. and Ullman, J. Adequacy of decompositions of relational database.

(8) Sagiv, Y. An algorithm for infering multivalued dependencies that warks also for a subset of propositional logic. J. ACM 27.2 (Apr. 1980), 250-262.

(9) Ullman, J. D. Principles of Database Systems. Computer Science Press, 1980.