関係データベーススキーマにおける関数従属関係のブール方程式を用いる取り扱いについて

竹島卓　富士通国際情報社会科学研究所

1. まえがき

本論文ではリレーショナルデータベース（RDB）における関数従属関係（FD）をブール代数によって取扱う方法について考察した．とくに従来の方法では困難であった「FDがない」という命題をもブール代数によって一貫して合理的に取扱えることを指摘した。ここに示した方法はFDばかりでなく，ある種の述語論理における推論をブール代数によって代数的に取り扱うことができ，応用範囲の広いものである。

RDBでは関係とよばれる，抽象的なデータ要素の集りを（一般に）複数個取扱う．このRDBに対する検索・追加・修正・削除の操作が簡単に誤りなく遂行でき，しかも関係に対応する記憶域の所要量も小さくて済むような関係の集合を求めることが重要である．このために，正規と呼ばれるいくつかの条件を満足するような関係の集合を求めることが提案されている．よく知られているCoddの才２正規形（２NF）や才３正規形（３NF）[1]，Boyce-Codd正規形[2]などは，FDの概念を基に定義されており，そのためにFDの性質を研究することが重要な意味をもっている．Delobel & Casey[3]はFDに論理関数を対応させ，この論理関数の最小表現を用いることによって正規化された関係の数を最小にすることを論じている．この後Bernstein[4]や上林[6]がこの方法を発展させ，欠陥を補って等価なkeyを含む場合のアルゴリズムを展開している．一方，この間にArmstrong[5]がFDについての研究を行って，その公理化を行っている．Fagin[7]はFDに対して含意命題を対応させてFDをとり扱う方法について論じている．しかしこれらの論理関数や含意命題にFDを対応させる方法は，便宜的であって，取扱いに注意しなければ誤った解釈をする可能性がある．また，「FDがない」という命題は対応する論理関数や含意命題をもたず，その情報を積極的に利用できないという欠点をもっている．有名な3NFでは推移従属という概念が用いられるが，これには「FDがない」ということが用いられており，（２NFでも「FDがない」は使われる．）しかも，「FDがない」ことの取扱いには多少のぎこちなさがみられ、一貫して合理的に扱う必要がある．現在ではFDを一般化した多値従属（multi valued dependency）がFagin[11]によって提案され，より複雑なdependencyの構造が考察されており，また増永[9]による関係間従属や関係内従属，Rissanen[8]のindependent componentなどの研究方向もある．本論文では，古典的なFDについて，その性質と取扱いについてテーマを絞り厳密に分析してゆく．

2. 基本的事項

ここではRDBにおける関係とそれに関連して用いらる諸概念で，FDを定義するために必要なものを簡単に説明する．以下では自然数nに対して$\bar{n}$で集合$\{1, 2, \cdots n\}$をあらわす．

まず，属性と呼ばれ互に区別される有限個の座標$A_1, A_2, \cdots, A_N$ $(N>0)$とそれら属性の各々に付随して，ドメイン（domain）と呼ばれる集合達$D_1, D_2, \cdots, D_N$とを考える．（A_iにはD_iが対応するものとする。）特別の元Λがあって，

どのドメインにも含まれる。 $\Lambda \in D_i$ (for $\forall i \in \bar{N}$)。 属性とドメインの値との順序対の集合で，次のもの，$t = \{<A_1,d_1>, <A_2,d_2>, \cdots, <A_N,d_N>\}$ （ただし $d_i \in D_i$ for $\forall i \in \bar{N}$.） を（座標系を$\mathcal{A}$とする）直積空間 $<\mathcal{A}, \mathcal{D}> = <\mathcal{A}, D_1 \times D_2 \times \cdots \times D_N>$ 上の tuple と呼ぶ。 ここに $\mathcal{A} = \{A_1, A_2, \cdots, A_N\}$ とする。 また $<x,y>$ は x と y とから成る順序対をあらわす。 tuple t の A_i 成分を $t.A_i$ であらわし，$t.A_i \triangleq d_i$ で定義する。 $\mathcal{B} \subset \mathcal{A}$ のとき t の $\mathcal{B}$ 方向への射影を $t.\mathcal{B}$ であらわし，$t.\mathcal{B} \triangleq \{e_1, e_2, \cdots, e_N\}$ で定義する。 ここに $A_i \in \mathcal{B}$ ならば $e_i = <A_i, d_i>$ とし，さもなくば $e_i = <A_i, \Lambda>$ とする。 $\mathcal{A}$ 上の1以上有限個の tuple から成る集合 $\{t_1, t_2, \cdots, t_M\}$ $(M>1)$ を$\mathcal{A}$上の関係Rと呼び Rel($\mathcal{A}$, R) であらわす。 直感的には Rel($\mathcal{A}$, R) は，直積 $\mathcal{D} = D_1 \times D_2 \times \cdots \times D_N$ のつくる空間の空でない有限部分集合であり，tuple $\{<A_1,\Lambda>, <A_2,\Lambda>, \cdots, <A_N,\Lambda>\}$ はその空間の原点に相当する。 本論文では部分空間は考えず全て $<\mathcal{A},\mathcal{D}>$ 上で考える。 $\mathcal{B}\subset\mathcal{A}$ のとき，関係Rの$\mathcal{B}$方向への射影を$R.\mathcal{B}$であらわし，$R.\mathcal{B} \triangleq \{t.\mathcal{B} \mid t \in R\}$ によって定義する。 記号・が tuple の成分と射影および関係の射影に同様に使用されているので注意してほしい。 本論文では関係内のFDの構造と性質とを扱うために $<\mathcal{A},\mathcal{D}>$ の部分空間の概念を定義しないが，そのことは，FDの本質にとって何らの障碍にもならない。 関係の分解と合成については稿を改めて述べるつもりである。

3. 関数従属（FD）

一般にFDといっているものには2種類ある．それらは$\mathcal{A}$上の関係Rについて言われるものと，$\mathcal{A}$上の関係の族 $\mathcal{R} = \{R_1, R_2, \cdots, R_M\}$ $(M \geq 1)$ についていわれるものとである． この区別は「FDがない」という概念を用いるとき重要となる． 今一つの混乱のもとは，FDを命題としたとき，それが 真 であるということと，命題の形式そのものとが区別がつきにくいことである．（命題の真偽はそこにあらわれる自由変数に定数を代入したときに決まる．）

［定義1］ $\mathcal{A}$上の関係Rと，X，Y$(\subset\mathcal{A})$とについてのR上のFD $X \xrightarrow{R} Y$ とは，次の命題のことである． $\forall t_1 \forall t_2 [(t_1 \in R) \wedge (t_2 \in R) \supset (t_1.X = t_2.X \supset t_1.Y = t_2.Y)]$．∎

すなわち「Rに属するどの1組の tuples についても，それらのX方向の射影が等しいならば，Y方向の射影も等しい。」という命題である．

［定義2］ $\mathcal{A}$上の関係の空でない有限の族$\mathcal{R} = \{R_1, R_2, \cdots, R_M\}$ $(M \geq 1)$ と属性集合XとYにについて$\mathcal{R}$上のFD $X \xrightarrow{\mathcal{R}} Y$ とは 命題 $\forall R[R \in \mathcal{R} \supset X \xrightarrow{R} Y]$ のことである．∎

これら命題の真偽はRや$\mathcal{R}$，X，Yに実際の値を代入したとき定まる． 次に「FDがない」に対する命題を定義しよう． これをNFDと呼ぶ．

［定義3］ $\mathcal{A}$上の関係Rと，属性集合X, Y$(\subset\mathcal{A})$ とについて NFD $X \xrightarrow{R} Y$ とは $\neg(X \xrightarrow{R} Y)$ のことである．∎

［定義4］ $\mathcal{A}$上の関係の族 $\mathcal{R} = \{R_1, R_2, \cdots, R_M\}$ $(M \geq 1)$ と属性集合 X, Y$(\subset\mathcal{A})$ とについて NFD $X \xrightarrow{\mathcal{R}} Y$ とは $\neg(X \xrightarrow{\mathcal{R}} Y)$ のことである．∎

後に明らかにするがR上（あるいは$\mathcal{R}$上）のあるFDが真であることは，何ら他のR($\mathcal{R}$)上のFDが真であることに影響を与えないが，NFDとは相容れないことがある．

$\mathcal{A}$上の形式的に存在し得るFD（あるいはNFD）連の個数は全部で 2^{2N} である． ある特定のR

については、その上で真となるFDの数とNFDの数との和は丁度2^{2N}であることも当然であろう。次にFDの閉包という概念を定義する。

［定義5］ R($\mathcal{R}$)上のFD達の集合（これを弱いFDschemaと呼ぶ）$\mathcal{F}=\{F_1, F_2, \cdots, F_m\}$(注)（$0 \leq m \leq 2^{2N}$）の閉包$\mathcal{F}^*$とは、$\mathcal{F}$のFD達が全て真であると仮定したときやはり真となるようなR($\mathcal{R}$)上のFD達の全てから成るFD集合のことである。すなわち $\mathcal{F}^*=\{F \mid \mathcal{F} \Rightarrow F$ となるFD$\}$。ここに$\Rightarrow$は推論できることを意味する。換言すると、R($\mathcal{R}$)に特定の関係を割当てることなしに述語論理の演繹を行うことである。$\mathcal{F}$が与えられたということは射影をとる属性集合達には、各々特定の属性集合が割当てられていると考える。▯

(注) $m=0$のときは空集合$\{\}$とする。

4. FDとブール代数

(4.1) PB変換

本項では、1変数述語論理をブール代数で扱う方法について述べる。

一般にある定まった定義域集合S($\neq \{\}$)と述語$P(x)$とがあるとき、Sの部分集合 $\{x \mid x \in S \wedge P(x)\}$ を、Sを母集合とする述語$P(x)$の指標集合と呼び、$Ind_1(P, S)$であらわす。Pに対応させた小文字pであらわすこともある。Sの部分集合の全体は和を$\cup$、積を$\cap$、補元を補集合、最大元をS、最小元を空集合とすることによって、ブール代数となる。このとき、述語命題$\forall x [x \in S \supset P(x)]$は$p=1$と同値になることが知られている。命題の中がもう少し複雑な（論理演算を含む）式になっている場合のために、命題－ブール式変換(PB変換)を定義する。

［定義6］ PB変換

$\mathcal{P}$が有限個の基本述語$P_1(x), P_2(x), \cdots, P_n(x)$および真値V、偽値Λとから論理演算子$\wedge$, $\vee$, $\neg$の演算を有限回適用して得られる命題とする。$\mathcal{P}$の0型PB変換$\mathcal{P}^{\#}$と1型PB変換$\mathcal{P}^{\flat}$とを次の6個の規則で帰納的に定義する。

(i) $\mathcal{P}$が真値Vのとき $\mathcal{P}^{\#}=0$, $\mathcal{P}^{\flat}=1$,
(ii) 偽値Λのとき $\mathcal{P}^{\#}=1$, $\mathcal{P}^{\flat}=0$,
(iii) 基本述語$P(x)$のとき $\mathcal{P}^{\#}=p'$, $\mathcal{P}^{\flat}=p$,
(iv) $\neg Q$のとき $\mathcal{P}^{\#}=Q^{\#}$, $\mathcal{P}^{\flat}=(Q^{\flat})'$,
(v) $Q \wedge R$のとき $\mathcal{P}^{\#}=(Q^{\#})'+(R^{\#})'$, $\mathcal{P}^{\flat}=Q^{\flat} \cdot R^{\flat}$,
(vi) $Q \vee R$のとき $\mathcal{P}^{\#}=(Q^{\#})' \cdot (R^{\#})'$, $\mathcal{P}^{\flat}=Q^{\flat}+R^{\flat}$. ▯

この定義のもとで

$$\forall x [x \in S \supset \mathcal{P}] \quad \text{は} \quad \mathcal{P}^{\#}=0 \quad \text{あるいは} \quad \mathcal{P}^{\flat}=1 \qquad (4.1\text{-}1)$$

と同値になる。

次に、命題が $\forall S [S \in \mathcal{S} \supset \forall x [x \in S \supset \mathcal{P}]]$ の形のものを扱うために、指標集合をこの型のもの用に再定義する。新しい指標集合$Ind_2(P, \mathcal{S})$は、

$Ind_2(P, \mathcal{S}) \triangleq \{x \mid x \in \mathcal{M} \wedge P(x)\}$ ただし、$\mathcal{M}=\bigcup_{S \in \mathcal{S}} S$ と定義する。この場合のブール代数としては$\mathcal{M}$の部分集合全体の集合算を考える。これによって

$$\forall S [S \in \mathcal{S} \supset \forall x [x \in S \supset \mathcal{P}]] \quad \text{は} \quad \mathcal{P}^{\#}=0 \quad \text{あるいは} \quad \mathcal{P}^{\flat}=1 \qquad (4.1\text{-}2)$$

と同値になる。

(4.2) FD/NFD－ブール式変換

本項では、前項での考察に基きFDやNFDをブール代数で扱う方法を述べる。まず、$\mathcal{A}$の各属性Aに対して述語$a(t)$を$t_1.A=t_2.A$で定義する。

ここに $\boldsymbol{t} \triangleq \langle t_1, t_2 \rangle$ とする。 すると空でない $X\ (\subset \mathcal{A})$ が $\{A_{i_1}, A_{i_2}, \cdots, A_{i_m}\}$ であるとき、述語 $t_1.X = t_2.X$ は $a_{i_1}(\boldsymbol{t}) \wedge a_{i_2}(\boldsymbol{t}) \wedge \cdots \wedge a_{i_m}(\boldsymbol{t})$ と同値になる。空集合 $\phi = \{\}$ に対しては $t_1.\phi = t_2.\phi$ は tuple の射影の定義より真値 $\curlyvee$ と同値である。 このことを用いて FD および NFD の定義を同値なものに書きかえておく。

$$(X \overset{R}{\to} Y) \equiv \forall \boldsymbol{t}\,[\boldsymbol{t} \in R \times R \supset (x(\boldsymbol{t}) \supset y(\boldsymbol{t}))] \qquad (4.2\text{-}1a)$$

$$(X \overset{R}{\not\to} Y) \equiv \neg \forall \boldsymbol{t}\,[\boldsymbol{t} \in R \times R \supset (x(\boldsymbol{t}) \supset y(\boldsymbol{t}))] \qquad (4.2\text{-}1b)$$

$$(X \overset{\mathcal{R}}{\to} Y) \equiv \forall R\,[R \in \mathcal{R} \supset \forall \boldsymbol{t}\,[\boldsymbol{t} \in R \times R \supset (x(\boldsymbol{t}) \supset y(\boldsymbol{t}))]] \qquad (4.2\text{-}2a)$$

$$(X \overset{\mathcal{R}}{\not\to} Y) \equiv \neg \forall R\,[R \in \mathcal{R} \supset \forall \boldsymbol{t}\,[\boldsymbol{t} \in R \times R \supset (x(\boldsymbol{t}) \supset y(\boldsymbol{t}))]] \qquad (4.2\text{-}2b)$$

ただし $X = \{A_{i_1}, A_{i_2}, \cdots, A_{i_m}\}\ (m \geq 1)$ のときは $x(\boldsymbol{t}) = a_{i_1}(\boldsymbol{t}) \wedge a_{i_2}(\boldsymbol{t}) \wedge \cdots \wedge a_{i_m}(\boldsymbol{t})$

$X = \{\}$ のときは $x(\boldsymbol{t}) = \curlyvee$

とする。 Y についても同様。

これらはいずれも前項で述べた PB 変換が可能な形をしている。 (4.2-1a) と (4.2-2a) とは 0 型 PB 変換のもとで いずれも

$$x y' = 0 \qquad (\cdot\ \text{記号は常に省略可能}) \qquad (4.2\text{-}3a)$$

となり、 (4.2-1b) と (4.2-2b) とは いずれも $\neg(xy'=0)$ すなわち

$$x y' \neq 0 \qquad (4.2\text{-}3b)$$

となる。 ただし $x(\boldsymbol{t}) = a_{i_1}(\boldsymbol{t}) \wedge a_{i_2}(\boldsymbol{t}) \wedge \cdots \wedge a_{i_m}(\boldsymbol{t})\ (m \geq 1)$ のときは

$x = a_{i_1} \cdot a_{i_2} \cdot \cdots \cdot a_{i_m}$

$x(\boldsymbol{t}) = \curlyvee$ のときは $x = 1$ とする。(y についても同様)

述語とその指標集合とを特に記号を変えなかった。 また $x \supset y$ は $x' + y$ とみなして PB 変換を行った。

[定義 7] (4.2-1a) ～ (4.2-2b) の FD や NFD を (4.2-3a) や (4.2-3b) に対応させる変換を、0 型の FD/NFD-ブール式変換(0 型 FB 変換)と呼び、もとの FD や NFD を F とするとき $F^{\#}$ であらわす。 特に F の PB 変換の対象となっている部分に注目するとき、その部分の PB 変換を F の指標関数と呼ぶ。 ▯

すなわち $X \to Y$ の指標関数 とはブール式 xy' であり、 $X \to Y$ の FB 変換とは $xy' = 0$ のことである。 F の指標関数を $\mathcal{I}(F)$ や $\hat{F}$ であらわす。

この指標関数のブール式(あるいはブール関数)が Delobel & Casey のいう論理関数である。 また、PB 変換を 1 型で行いその結果において $x' + y$ を $x \to y$ でおきかえれば、これが丁度(大文字と小文字の違いはあるが) Fagin の含意命題になっている。 論理関数にせよ含意命題にせよ、2 値であるために、「FD がない」に対応するものが考えられない。 また含意命題というのはみかけだけで命題の意味はもっていないため、推論規則も制限される。 これに対して、本項で導入した FB 変換は、もとの FD の意味を完全に含んでおり、ブール代数の演算と通常の推論が自由に行える点が優れている。

(4.3) ブール代数の諸定義と諸定理

本項では、ブール代数の上で推論を行う際に有用な諸定理を示す。

[定義 8] ブール関数 $f(x_1, x_2, \cdots, x_n)$ の resultant $r(x_{\sigma(1)}, x_{\sigma(2)}, \cdots, x_{\sigma(m)}) = f(x_{\sigma(1)}, x_{\sigma(2)}, \cdots, x_{\sigma(m)} \mid x_{\sigma(m+1)}, \cdots, x_{\sigma(n)})$ $\left\{\begin{array}{l}(\sigma \text{ は } 1, 2, \cdots, n \text{ の置換}) \\ (0 \leq m \leq n)\end{array}\right\}$ は次のように定義される。

$$r(x_{\sigma(1)}, x_{\sigma(2)}, \cdots, x_{\sigma(m)}) \triangleq \prod_{\langle t_1, t_2, \cdots, t_\ell \rangle \in B^\ell} f(x_1, x_2, \cdots, x_n) \Big|_{\langle x_{\sigma(m+1)}, x_{\sigma(m+2)}, \cdots, x_{\sigma(n)} \rangle = \langle t_1, t_2, \cdots, t_\ell \rangle}$$

ここに　$B^{\ell}=\{<b_1,b_2,\cdots,b_\ell> \mid b_1,b_2,\cdots,b_\ell\in\{0,1\}\}$，　$\ell=n-m\,;\,n,m,\ell\geq 0$　である．
∎

[定理 1]　(resultant 定理)
$f(x_1,x_2,\cdots,x_n)$ をブール関数，$r(x_{\sigma(1)},x_{\sigma(2)},\cdots,x_{\sigma(m)})$ を f の任意の resultant とすれば，

$$\exists x_1 x_2\cdots x_n\ f(x_1,x_2,\cdots,x_n)=0 \iff \exists x_{\sigma(1)}x_{\sigma(2)}\cdots x_{\sigma(m)}\ r(x_{\sigma(1)},x_{\sigma(2)},\cdots,x_{\sigma(m)})=0$$

が成立つ．∎　($\exists x_1x_2\cdots x_n$ は $\exists x_1\exists x_2\cdots\exists x_n$ の略記．$\forall$ についても同様)

[系 1]　ブール関数 f について $f=0$ が有解である必要かつ十分条件はその 0 変数 resultant が 0 となることである．∎

[定理 2]　(解の公式)
$f(x_1,x_2,\cdots,x_n)=0$　が有解であるならば，解は

$$x_1^*=f(0,x_2^*,x_3^*,\cdots,x_n^*)+u_1\cdot f'(1,x_2^*,x_3^*,\cdots,x_n^*)$$
$$x_2^*=f(0,x_3^*,\cdots\ x_n^*\mid x_1)+u_2\cdot f'(1,x_3^*,\cdots,x_n^*\mid x_1)$$

$$x_n^*=f(0\mid x_1,x_2,\cdots,x_{n-1})+u_n\cdot f'(1\mid x_1,x_2,\cdots,x_{n-1})$$

であたえられる．　ここに　$u_1,u_2,\cdots,u_n$ は任意定数である．　∎

[定理 3]　(不等式の解)
(3-i)　$xy=0\wedge y\neq 0 \implies x\neq 1$
(3-ii)　$\exists y(xy=0\wedge y\neq 0)\iff x\neq 1$　∎

[定理 4]　(resultant の最大性定理)
ブール関数 $f(x_1,x_2,\cdots,x_n)$ と $g(x_{\sigma(1)},x_{\sigma(2)},\cdots,x_{\sigma(m)})$　($m\leq n$ とする) および g と丁度同じ変数をもつ f の resultant $r(x_{\sigma(1)},x_{\sigma(2)},\cdots,x_{\sigma(m)})$ とについて，

$$\forall x_1\forall x_2\cdots\forall x_n[f=0\Rightarrow g=0]\implies\forall x_{\sigma(1)}\forall x_{\sigma(2)}\cdots\forall x_{\sigma(m)}[r=0\Rightarrow g=0]$$

が成立つ．∎

[定義 9]　ブール関数 f のすべての変数に 0 あるいは 1 を代入するとき，その任意の代入において f の関数値が常に 0 あるいは 1 であるような場合 f を整関数と呼ぶ．∎
[定義 10]　ブール変数あるいはその否定の積であって 0 でないものを積項と呼ぶ．∎
[定理 5]　$f(x_1,x_2,\cdots,x_k,x_{k+1},\cdots,x_n)$ を整関数，$g(x_1,x_2,\cdots,x_k)$ を積項とする．

$$\forall x_1x_2\cdots x_n[g=1\Rightarrow f=1]$$
$$\implies\forall x_1x_2\cdots x_n[f=0\Rightarrow g=0]\qquad ∎$$

[定理 6]　$f(x_1,x_2,\cdots,x_k,x_{k+1},\cdots x_n)$ は整関数，$g(x_1,x_2,\cdots,x_k)$ も整関数とする．
$f=0$ が有解かつ $g=1$ が有解とすれば

$$\forall x_1x_2\cdots x_k\,\exists x_{k+1}x_{k+2}\cdots x_n\,[g=1\wedge f\neq 1]$$
$$\implies\neg\forall x_1x_2\cdots x_n[f=0\Rightarrow g=0]\qquad ∎$$

(4.4) 弱いFD schemaと閉包について

本節では弱いFD schemaが与えられたときその閉包を求める問題を考える。弱いFD schema $\mathcal{F}$ を $\mathcal{F}=\{F_1, F_2, \dots, F_m\}$ としよう。$\mathcal{F}$ の0型schema-Boolean変換(SB変換) $\mathcal{F}^\#$ を $\mathcal{F}^\#=\{F_1^\#, F_2^\#, \dots, F_n^\#\}$ で定義すると、F_i の指標関数 f_i によって、$\mathcal{F}^\#=\{f_1=0, f_2=0, \dots, f_m=0\}$ とかける。これは連立方程式と解釈できるが、ブール代数においては $(x+y=0)$ と $(x=0$ かつ $y=0)$ とは同値であるから、$\mathcal{F}^\#$ は単一の方程式 $\Sigma_{i=1}^{m} f_i=0$ と同値である。この左辺を弱いFD schema $\mathcal{F}$ に対応する指標関数といい $\mathcal{I}(\mathcal{F})$ あるいは $\hat{\mathcal{F}}$ であらわす。

$\mathcal{F}$ と $\hat{\mathcal{F}}=0$ とは命題として同値($\mathcal{F}$ の各元を論理積で結んで考える。)であるから $\mathcal{F}$ から演繹できることは $\hat{\mathcal{F}}=0$ からも演繹できなければならない。むろん逆も成立つのだが、$\hat{\mathcal{F}}=0$ には、FDについてのみ適用することを考えれば、実際的に役立つ手続きを与える次の定理がある。

［定理7］

$$F\in\mathcal{F}^* \iff f=1 \supset \hat{\mathcal{F}}=0 \qquad \text{ここに } f=\mathcal{I}(F). \ \|$$

(証明) 定理5と定理6とを適用する。‖

この定理を用いれば全ての $\mathcal{A}$ 上のFD達から成る集合 $\mathcal{H}$ を $\mathcal{F}^*$ と $\mathcal{G}\triangleq\mathcal{H}-\mathcal{F}^*$ の2つに分解できる。一方Ⓡ(Rあるいは $\mathcal{R}$ の意味で用いる)を与えると、その上でFDの真偽が定まるからⓇによっても $\mathcal{H}$ の2分割 $\mathcal{H}=\mathcal{F}_{Ⓡ}\cup\mathcal{G}_{Ⓡ}$, $\mathcal{F}_{Ⓡ}\cap\mathcal{G}_{Ⓡ}=\{\}$ が定まる。ここに $\mathcal{F}_{Ⓡ}$ はⓇの上で真なるFDの集合, $\mathcal{G}_{Ⓡ}$ は偽なるFDの集合である。

Ⓡ上の弱schema $\mathcal{F}$ が成立つというのは $\mathcal{F}^*\subset\mathcal{F}_{Ⓡ}$ となることである。

この $\mathcal{F}^*$ と $\mathcal{F}_{Ⓡ}$ とが一致するような関係Rの存在がArmstrongによって証明されている。　　　　　。このことは後に強いFD schemaのところで述べる。

次に、$\mathcal{F}$ から $\mathcal{F}^*$ を全て求める簡単なアルゴリズムを述べる。これは次節でDelobel & Caseyの6公理の証明に用いる。

(アルゴリズムi)

FDを右辺が空集合のもの，単一属性の集合のもの，2属性以上の集合のものの3とおりに分ける。

(step 1) 任意のXについて $X\to\{\}$ は自明であるから $\mathcal{F}^*$ の元である。

(step 2) $\hat{\mathcal{F}}$ の積和展開で非零の係数をもち補元因子を唯1つ含む項に対応するFDは $\mathcal{F}^*$ の元である。

(step 3) $\hat{\mathcal{F}}$ の全てのresultant達にわたって(step 2)と同様のことを行う。

(step 4) step 3までに決定済のFD達について左辺が同じもの同士を次節の才1項(5.1)の基本的性質5°を適用してまとめる。‖

(アルゴリズムの正当性証明は省略)

5. ブール代数によるFDの解析

(5.1) 基本的性質

$\mathcal{A}$ を全属性集合, $X, Y, Z, W\subset\mathcal{A}$ とし, x, y, z, w を対応する指標集合とする。$A_i\in\mathcal{A}$ に対して a_i を対応する指標集合とする。(0型FB変換)

1° reflexivity $X\to X$ $\because\ xx'=0$.

2° augmentativity $X\to Y$ ならば $X\cup Z\to Y$ $\because\ xy'=0$ ならば $xzy'=0$.

3° pseudo-transitivity $X\to Y$ かつ $Y\cup Z\to W$ ならば $X\cup Z\to W$

$\because\ xy'=0$ かつ $yzw'=0$ ならば $xzw'=0$.

4° projectivity $X\cup Y\to Y$ $\because\ xyy'=0$.

5° additivity $X\to Y$ かつ $Z\to W$ ならば $X\cup Z\to Y\cup W$

$\because\ xy'=0$ かつ $zw'=0$ ならば $xz(yw)'=0$.

6° transitivity　$X\rightarrow Y$ かつ $Y\rightarrow Z$ ならば $X\rightarrow Z$ $\because\ xy'=0$ かつ $yz'=0$ ならば $xz'=0$

7°　$X\rightarrow\{\}$　$\because\ x1'=0$

8°　$\mathcal{A}\rightarrow X$　$\because\ ax'=a_1\cdot a_2\cdot\cdots\cdot a_N(a_{i_1}\cdot a_{i_2}\cdots a_{i_m})'=0\ (0\leq m\leq N)$

この最後の2つには名がついていない。空集合{}にブール代数の最大元1が対応していることに注意しておく。$\mathcal{A}$は必ずしも最小元0に対応しない。(この点、本学会18回大会の筆者の報告は誤りであった。ここに訂正しておく。)

1°～6° を Delobel & Casey の6公理と呼ぶ。実はこのうちの (1°, 2°, 3°) あるいは (4° 5° 6°) より他の性質は全て導かれることは Armstrong が指摘している。筆者は (1° 7° 3°) からも他の性質が全て導かれることを証明している。証明は 3°において $Y=\{\}$ としたものに7°を適用することによって2°が得られることを確かめればよい。

(5.2) FFD とブール関数の prime implicant .

[定義11] 弱い FD schema $\mathcal{F}$ が与えられたとき、$F\in\mathcal{F}^*$ である F について、FD $F(=X\rightarrow Y)$ が全関数従属 (FFD) であるというのは、Xが空でなく、かつ X の真部分集合Wのすべてについて $W\rightarrow Y\notin\mathcal{F}^*$ となることをいう。 ‖

[定理8] 右辺が単一の属性から成る FD Fが与えられた弱い FD schema $\mathcal{F}$ において FFD であるための必要かつ十分な条件は、$\mathcal{F}$ の指標関数 $\hat{f}$ に対して、F の指標関数 f が $\hat{f}$ の prime implicant となることである。‖

FFD は関係の key を定めたり、2NF, 3NF を定義するための重要な概念であり、関係の分解や合成の際、あるいは RDB に対する問合せのアクセスパス決定のための重要な情報であるが、本論文ではこれ以上立入らない。

(5.3) 弱い FD schema の無矛盾性と充足性.

弱い FD schema $\mathcal{F}$ と対応するブール方程式 $\hat{f}=0$ とが同値であることは已に述べた。$\hat{f}$ の形を考えると、それは $\sum_{i=1}^{n}(a_{i_1}\cdot a_{i_2}\cdot\cdots\cdot a_{i_m})(b_{i_1}\cdot b_{i_2}\cdot\cdots\cdot b_{i_l})'$ の形をしている。ゆえに、全ての変数を1にすれば $\hat{f}\big|_{全ての変数を1}=\sum_{i=1}^{n}(1\cdot 1')=0$ となるから $\hat{f}=0$ は必ず解をもつ。ゆえに、これに同値な $\mathcal{F}$ は無矛盾である。

[定理9] 弱い FD schema は無矛盾である。‖

さらに強く次の定理が成立つ.

[定理10] 任意の弱い FD schema を満足する関係R (や関係の族$\mathcal{R}$) が存在する.‖

(証明) $R_\Lambda=\{\{\langle A_1,\Lambda\rangle,\langle A_2,\Lambda\rangle,\cdots,\langle A_N,\Lambda\rangle\}\}$ や $\mathcal{R}_\Lambda=\{R_\Lambda\}$ は定理の条件にかなう。‖

(5.4) Delobel & Casey の6公理 (DC6A)

5.1項の基本的性質のうち (1°, 2°, …, 6°) を DC6A と呼ぶ. これらが FD の公理として完備であることはつぎのようにして確かめられる.

アルゴリズム i によれば $\mathcal{F}^*$を求めるために先ず $X\rightarrow\{\}$ すなわち 7°を用い、step 4. で 5°を用いた。これ以外に、積和展開をつくることと、resultant をとることとが用いられる。この手続きが DC6A の性質のみを用いて構成できることを確認する。7°は 4°において $Y=\{\}$ という特別の場合であるから問題はない。5°はもちろん問題ない。積和展開は展開した各項のうち非零でしかも補元因子を唯1つ含むもののみを求める為に用いるのだから、結果的には、もとの各項に補元因子でない変数を追加するためだけに用いたことになる。したがって、2°の augmentativity の適用によって可能である。最後に resultant をとるこ

とが良く知られる consensus 演算を全ての項に対して同時に行うことと同等であることを確かめるならば，これが3°の pseudo-transitivity によって可能であることが了解される．

6. 強い FD schema の初歩的考察

(6.1) 強い FD schema のブール代数

強い FD schema とは FD の schema $\mathcal{F}$ と NFD の schema $\mathcal{G}$ の2つを組にしたものである．NFD の schema とは NFD の集合のことである．強い FD schema を $\langle \mathcal{F}, \mathcal{G} \rangle$ であらわす．4.2項で述べた FD/NFD の FB変換より $\langle \mathcal{F}, \mathcal{G} \rangle$ は，次の連立方程式（不等式も含めて）と同値である．

$$f_1=0,\ f_2=0,\ \dots,\ f_m=0, \qquad (6.1-1)$$

$$g_1\neq 0,\ g_2\neq 0,\ \dots,\ g_n\neq 0. \qquad (6.1-2)$$

ここに f_i は $F_i(\in \mathcal{F})$ の指標関数，g_j は NFD G_j の指標関数で，NFD G の指標関数とは $\neg G$ なる FD の指標関数のこととする．

不等式の解の公式によって，slack 変数 $u_1, u_2, \dots, u_n$ を導入すれば (6.1-2) は

$$u_1 g_1'=0,\ u_2 g_2'=0,\ \dots,\ u_n g_n'=0,\ \text{かつ} \qquad (6.1-3)$$

$$u_1\neq 0,\ u_2\neq 0,\ \dots,\ u_n\neq 0 \qquad (6.1-4)$$

と同値になる．

したがって (6.1-1) および (6.1-2) の連立方程式が有解であることは，連立方程式 (6.1-1) および (6.1-3), (6.1-4) が有解であることと同値になる．

(6.1-1) と (6.1-3) とをまとめると，

$$\sum_{i=1}^{m} f_i + \sum_{j=1}^{n} u_j g_j' = 0 \qquad (6.1-5)$$

となる．

(6.1-5) の左辺の関数の resultant で $u_1, u_2, \dots, u_n$ 以外の変数を消去した関数を $\psi(u_1, u_2, \dots, u_n)$ とすると，(6.1-1) と (6.1-2) の有解性は，$\psi=0$ とおいたとき，どの u_k も非0解をもつことと同値である。以上より次の定理を得る．

［定理11］

強い FD schema $\langle \mathcal{F}, \mathcal{G} \rangle$ は，$\mathcal{F}$ と $\mathcal{G}$ に対応するブール方程式（不等式を含む）に slack 変数を導入して作った関数 $\sum f_i + \sum u_j g_j'$ の $u_1, u_2, \dots, u_n$ を残した resultant を0とおいた方程式が，すべての u_k について非0解をもつとき，かつそのときに限り無矛盾である．□

(6.2) 強い FD schema 上の推論

強い FD schema 上の推論は等式のみの弱い FD schema のように容易なわけにはゆかない．むろん FD の部分の推論は弱い FD schema と同様に可能である．しかしながら NFD に関する推論は FD の部分が必要であって，どうしても FD 部分と NFD 部分の両方を組合せた (6.1-5) のような式を扱う必要がある．

簡単な例をやってみよう．

$A\to B,\ B\to C,\ B\not\to A$ が与えられたとする．対応する式は，

$f(a,b,c,u)=ab'+bc'+u(ba')'=0,\ u\neq 0,$

となる．無矛盾性をみるために u を残す resultant をとってみよう．結果は $\psi(u)\equiv 0$（恒等0）となりしたがって u は任意の解をもち，特に $u\neq 0$ の解をもつ．ゆえに上例の強い FD schema は無矛盾である．

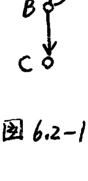

図6.2-1a

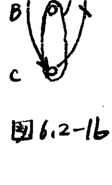

図6.2-1b.

つぎに，uを消去して $f(a,b,c|u)=ab'+bc'$ これは FD 部分をとりだしたことになっている．これから $ab'=0$, $bc'=0$ ，$ac'=0$(注)で新しい FD $\{A\}\to\{C\}$ が得られる．

$f(a,b,u|c)=ab'+ub'+au$ これから $ua=0$, $ub'=0$ …（EX-1）．

$f(a,c,u|b)=ac'+uc'+au$ これから $uc'=0$ （EX-2）．

$f(b,c,u|a)=bc'+ub'$ これからは新しい情報は特に得られない．

（EX-1）と（EX-2）とから $u\neq0$ という条件を使って $a'\neq0$, $b\neq0$, $c\neq0$

これは，A にはどこからも（Aを含むもの以外には）FD がない，つまり，FDの頂点にいることを示している．$b\neq0$ と $c\neq0$ は BとCとが決して key になれないことを示している．さらに（EX-1）と（EX-2）とを組合せると，$u(a+b'+c')=0$ となるが，これから $bca'\neq0$ つまり$\{B,C\}\not\to\{A\}$が得られる．この他に $u(a+b')=0$, $u(a+c')=0$ から $\{B\}\not\to\{A\}$ と$\{C\}\not\to\{A\}$とが得られる．

これを図示したのが図6.2-1a と図6.2-1b である．しかし，このような簡単な例でも，与えられたFD/NFD の強schema から推論できるNFD をすべて求めることはむつかしい．（見落しがないかどうかわからない．）この点強いFD-schema においては，閉包のような概念は考えにくい．というのは一般に強い schema のブール方程式を変形して slack 変数を1つだけ含み他の slack 変数でない変数のすべてを含む resultant $r(x_1,x_2,\cdots,x_N,u)$を得たとする．これから一般に（$r=0$ とおいて）$u\cdot\prod_{j\in J}(x_j')=0$ という uといくつかの補元変数でない変数との積が0であることが導ける．この他に $u\cdot x_k'=0$ （$k\notin J$）が導かれる可能性があるが，これと合わせて $u(x_k'+\prod_{j\in J}x_j)=0$ が得られ$u\neq0$ を用いると $\sum_{j\in J}(x_k\cdot x_j')\neq0$ となる．これは FD でいえば $\neg\{(X_k\to X_{j_1})\wedge(X_k\to X_{j_2})\wedge\cdots\wedge(X_k\to X_{j_{|J|}})\}$ となり，否定を中に入れて $(X_k\not\to X_{j_1})\vee(X_k\not\to X_{j_2})\vee\cdots\vee(X_k\not\to X_{j_{|J|}})$ となるが，これを自然語に訳すと，「X_k から X_{k_1} への FD がないか，X_kからX_{j_2}への FD がないか，…，X_kから$X_{j_{|J|}}$ への FD がないかである。」となって，はなはだ難解である。

(注) 上から2行目．前の2式からbを消去した．

(6.3) Armstrong の才5定理．

強い FD schema の無矛盾性は 定理11 によって検定できるが，無矛盾な場合の充足性（R あるいは$\mathcal{R}$ の存在）については主張できない．Armstrong は，閉包$\mathcal{F}^*$ とそれに属さない FD を全て NFD として作った集合 $\mathcal{G}^*$ からなる強い FD schema $\langle\mathcal{F}^*,\mathcal{G}^*\rangle$ について次の定理を証明した．

［定理12］ 強い FD schema（上述のように定義される）$\langle\mathcal{F}^*,\mathcal{G}^*\rangle$ を満足する関係Rが存在する．（Armstrong）‖（出典の定理とは異るが同等である。）

この定理を用いれば容易に次の定理が導かれる．

［定理13］ 任意の無矛盾な強い FD schema $\langle\mathcal{F},\mathcal{G}\rangle$ を満足する関係R（あるいは$\mathcal{R}$）が存在する． ‖

証明は$\mathcal{F}^*$ が$\mathcal{G}$に矛盾しないことから定理12の場合に帰着できることから明らかである．

7. 結論と考察

本論文の要点は次の三点である．

才一に，リレーショナルデータベースにおける関数従属関係（FD）について，重要であるにもかかわらず従来は軽視されていた「FD がない。」という概念を取り扱う上で，一つの有用な方法を示したこと．

才二に，FD が（ないというも含めて）完全にブール代数で解釈可能であるこ

とを理論的に証明したこと．
　オ三は，このようなブール代数において，resultant 定理といわれる強力で美しい定理があり，それが有効に利用できることを確認したこと．
　以上の点である．
　とくにオ一の点に関しては，従来 FD の理論的研究で用いられている2値の論理代数や命題算によってはとり扱い得ないことを指摘しておこう．　オ二の点に関しては，充足可能性（というより実際の関係や関係の族の存在）の証明が Armstrong のオ5定理に頼っている点が不満である．　オ三の点に関しては，必ずしも2値でないブール代数に還元できる諸問題，たとえば1変数述語論理の問題など，がうまく解釈できることを期待できる．　とくに resultant 定理を使用して，ブール方程式の一般解の解表示を示したものは，いくつか調査した文献には見出し得なかった．　筆者がこの定理を知ったのは1966年頃の数学セミナー誌上で，故 山田欽一 一橋大教授の書かれたものを読んだときである．　教授は，この定理を用いて，一階一変数の述語論理の推論を行う機械 syllogister を作成なさったとのことである．　筆者は現在，ブール代数式記号処理システムを作成中であるが，山田教授にちなんで，これを syllogister と命名した。
　本論文では触れ得なかったが，FD を用いた関係の分解や合成の理論は重要である．　この問題については，さらに研究の上報告する予定である．

謝辞　本論文は国際情報社会科学研究所のデータベースプロジェクトで研究の発端を得たものである．　日頃御指導いただいている北川敏男所長をはじめ，研究所の諸氏に感謝いたします．　とくに，熱心に討論をし，貴重な示唆をしていただいた 国藤進氏 には深く感謝の意を表します。

文献

1. Codd, E. F. : Relational completeness of data base sublanguages, Data Base Systems, 1971.
2. Codd, E. F. : Further normalization of data base relational model, Data Base Systems, 1971.
3. Delobel, C. & Casey, R. G. : Decomposition of a data base and the theory of Boolean switching functions, IBM J. Res. Develop, 17-5, 1972, pp 374-386.
4 Bernstein, P.A. : Synthesizing third normal form relations from functional dependencies, TODS, 1-4, 1976, pp 277-298.
5 Armstrong, W.W. : Dependency Structures of data base relationships, Inf. Proc. 74, 1977, pp 580-583,
6 上林弥彦 ：リレーショナル データベースの最適な正規表分解について，研究会資料DB33-3
7 Fagin, R : Functional Dependencies in a Relational Data Base and Propositional Logic, IBM Res. Rep. RJ1976, IBM J. RES. DEVELOP. に掲載されているはず。
8 Rissanen, J : Independent Components of Relations, TODS, vol.2, 1977, pp 317-325.
9 増永良文，野口正一：関係间関数依存性を考慮した関係データベースの一構成法，信学技報 AL 77-36, 1977, pp 93-99.
10 山田欽一： カッコいい代数（ブール代数とその応用），数学セミナー誌，1966年頃．古い文献なので出典未詳・副題もあまり確かではない．『1977, pp 262-278
11 Fagin, R : Multi-valued dependencies and a new normal form for relational data-bases, TODS, vol.2,