

汎関係データベース・モデル

データ・モデル, 関係演算系, スキーマの汎化について

国藤 進, 小林 要

(富士通・国際情報社会科学研究所)

〔概要〕 関係データベースの新しい概念モデルとして汎関係というデータベース・モデルを提案し, データ抽象化という見地からの汎関係の定式化を行い, 汎関係上での関係演算系の性質を論じ, 条件付従属という新概念を用いスキーマの汎化について考察した。

1. はじめに

1970年にCoddが関係モデル(RM)を提唱し, 6)して以来, Coddモデルに対して種々の議論が投げかけられてきたが, それらは主として, 次の二側面からなされた。すなわち,

(1) relational level \longrightarrow data specification(2) access-path level \longrightarrow data realization

である。前者に関してはデータ抽象化の問題が, 後者に関しては正規化の問題が, 特に活発に議論されている。Liskov⁵⁾が, 高水準言語におけるデータ抽象化の必要性を喚起してから, 関係データベース(RDB)の世界にもデータのもつ意味論的世界を反映しやすいデータ構造を導入すべきだという意見が強まり, 最近Smith³⁾は, aggregation 抽象化と generalization 抽象化という注目すべきデータ抽象化の存在を指摘した。前者はHoare⁴⁾の直積型データ構造, 後者は直和型データ構造に対応するが, それらを用いて直接に対象を表現しようとしたため, 階層的構成の仕方がかなり複雑となった。これら抽象化の要請の本質を洞察するに, 対象から特定の視点で切り取られた実在たるRDBのもつあらゆる条件を, 明白に自然な形態で表現するにはいかにすべきかが問われているのに着目した。他方, いわゆる正規化問題についてはSchmid⁵⁾の批判以来, 多くの研究者によつて議論されている。関数従属(FD)に関するDeLoob⁶⁾, Bernstein⁷⁾, Armstrong⁸⁾, 上林¹⁰⁾, Rissanen¹¹⁾, 増永¹²⁾, 竹島³³⁾の研究, 多値従属(MVD)に関するFagin¹³⁾, 上林¹⁶⁾, 牧久内¹⁷⁾の研究が注目される。なかでも従属関係の理論の中に, 場(context)の問題を持ち込んだFagin, 上林のMVD研究は, 一考に値する。これら従属関係の理論の立場から, やはり条件付されたFD・MVDという概念を明らかにする必要性を喚起された。

上述の諸議論を踏まえて, われわれは複数のCoddの関係性, 単一の関係性に埋込む汎関係という新しいDBモデルを構築した。汎関係とはnull値を導入することによつて複数の関係表を一枚の関係表のごとく見せるものであり, 汎関係化することによつて, 大規模DBの応用ユーザおよびDB設計者の双方にとつて, 平坦かつ柔軟な見方を提供できる。そのような種々の見地からのアクセスを可能にするため, われわれはタグ関係と条件付関係というデータ概念を導入した。タグ関係というのがSmith³⁾のいうgeneralizeする時のカテゴリーの軸に相当し, 特定のカテゴリーで切り取られた関係が条件付関係である。更にわれわれ

これは、スキーマの汎化に際して条件付FD, 条件付MVDという概念を定式化した。すなわち汎関係モデルの中心概念は、汎化および条件化である。

さて1962年にCODASYL¹⁹⁾により、提議され、小林¹⁷⁾により、拡張された情報代数モデルも、単一化された情報空間をもつという点では汎関係モデルと同じである。いわば汎関係モデルは情報代数モデルの"関係"版であるが、より自然な条件化演算によりANSI/SPARC²⁰⁾という概念スキーマ・外部スキーマ・内部スキーマ間の橋渡しを円滑に行い、また実際に実働化された棒のDPLS²⁰⁾, 弘原海²¹⁾のGEDAS²²⁾両システムの理論的ひな型としての機能をも果しうると考えられる。

2. データベース・モデル

2.1 null値の導入と汎化の概説

現実界を特定の人間(達)が特定の視座に従って切断し、われわれにとって理解可能なデータ構造上へ不完全ながら写像した時、われわれは多くの対象の反映としてのデータの塊り、すなわちデータ単体をもつ。このようなデータ単体を計算機にとって表現可能なデータ構造上へ写像しデータ表現していく訳であるが、関係モデルでは正規化された表形式(、すなわち第一正規形)の関係に、データ単体を写像していく。汎関係モデルでは複数の第一正規形の関係を、単一のフラットな第一正規形の関係へと汎化していく。この汎化過程を手続き的に示すのはやさしいが、ここでは直観的に例示する。

H ::=

H#	HName	Manu.	Price	HQual.
H ₁	A180	A	2000	25
H ₂	B160	B	1500	20
H ₃	B190	B	3000	30

SY ::=

H#	S#	SYQual.
H ₁	S ₁	200
H ₂	S ₁	180
H ₂	S ₂	210
H ₃	S ₁	250
H ₃	S ₂	280

S ::=

S#	SName	SQual.
S ₁	OSA	10
S ₂	OSB	20

H : Hardware
S : Software
SY : SYstem

図1. 関係H-S-SY

G ::=

RelName	H#	HName	Manu.	Price	HQual.	S#	SName	SQual.	SYQual.
H	H ₁	A180	A	2000	25	Λ	Λ	Λ	Λ
H	H ₂	B160	B	1500	20	Λ	Λ	Λ	Λ
H	H ₃	B190	B	3000	30	Λ	Λ	Λ	Λ
S	Λ	Λ	Λ	Λ	Λ	S ₁	OSA	10	Λ
S	Λ	Λ	Λ	Λ	Λ	S ₂	OSB	20	Λ
SY	H ₁	Λ	Λ	Λ	Λ	S ₁	Λ	Λ	200
SY	H ₂	Λ	Λ	Λ	Λ	S ₁	Λ	Λ	180
SY	H ₂	Λ	Λ	Λ	Λ	S ₂	Λ	Λ	210
SY	H ₃	Λ	Λ	Λ	Λ	S ₁	Λ	Λ	250
SY	H ₃	Λ	Λ	Λ	Λ	S ₂	Λ	Λ	280

G : Generalized relation

Λ = "null value"

図2. 汎関係G

例題用 RDB として、図 1 のような三つの関係 H, S, SY の実体が存在する
と考える。この時、これらを一の汎関係 G の中に埋込むと、図 2 のようにな
る。すなわち新しく Rel. Name という属性を仮定し、それに対応する関係の
(実体でなく) 名前 H, S, SY を放り込み、元の表形式データもあるがままに
放り込み、それ以外の所には今迄に使用されていない特殊記号 Λ (null 値) を割
当てるのが自然である。

データ $\left\{ \begin{array}{l} \text{既獲得} \\ \text{未獲得} \end{array} \right\} \left\{ \begin{array}{l} \text{獲得可能} \\ \text{獲得不可能} \end{array} \right\} \left\{ \begin{array}{l} \text{概念的定義不可能} \\ \text{概念的定義可能} \\ \text{概念的定義不可能} \end{array} \right\}$

図 3. null 値の多義性 (・印: null 値の候補)

汎関係における null 値 Λ
は、図 3 の・印のごとく種
々の解釈が可能である。
情報代数においても Θ (
Missing, relevant but
not known) と Ω (Unde-
fined, not relevant) の

区別が重要である。多義な Λ を解釈する問題は、汎関係 DB の利用法に依存す
るので、その利用時において Λ の多義性を識別し取扱うデータ操作言語 (dml)
) を提供すればよい。本小論では、 Λ の単一の解釈のみを許容するデータ定義
言語 (ddl)、 dml を構成することにする。

2. 2 汎関係の定式化

前節のモデルを正当化し、実存するデータを順序対 \langle 属性, 値 \rangle (この「値」
は、正確にいうと現実の世界の「量」²⁰⁾ である。) の集まりとみなす立場に徹し
、汎関係の定式化を行う。

<定義 1> 唯一の共通元が null 値 Λ のみであるような、互いに異なる空で
ない集合 Δ_i ($i = 1 \dots p$) の族を 基本型集合 (base type set) $\Delta \triangleq \{ \Delta_1, \dots,$
 $\Delta_p \}$ と呼ぶ。すなわち、

$$\forall i \forall j ((i, j \in \{1, \dots, p\}) \wedge (i \neq j) \Rightarrow \Delta_i \cap \Delta_j = \{ \Lambda \}).$$

ここに「量」の変換を行わない通常の Algol 系言語においては、 Δ_i の候補とし
て integer, real, boolean, string 等が考えられる。

<定義 2> 一意識別可能な属性 a_i ($i = 1 \dots N$) をもち、 Δ のある元を値
とするような集合 Θ_i を 定義域 と呼ぶ。

ここに N はありとあらゆる可能な属性を含むよう十分な大きな値をとることが要
請される。すなわち理論上は非可算無限個の属性を内包しえることが要請され
るが、実用上は応用ユーザ特有の視座で切断した属性を内包しえるだけの濃度が
あればよい。

<定義 3> 全属性集合 $\Omega \triangleq \{ a_1, \dots, a_N \}$ 上の属性識別子 α は、次のよう
に定義される:

$$\alpha \triangleq \left\{ \begin{array}{l} (a_{i_1}, a_{i_2}, \dots, a_{i_n}) \text{ ただし } 1 \leq n \leq N, a_{i_j} \in \Omega (j \in \{1, \dots, n\}, \\ i_j \in \{1, \dots, N\}, (i \neq i' \Rightarrow a_{i_j} \neq a_{i'_j})), \\ \text{() 最もしくは。 (空属性識別子)} \end{array} \right.$$

<定義 4> 定義域 $\Theta_{i_1}, \dots, \Theta_{i_n}$ 上の n 項直積, n 項関係, n 項組はそれ
ぞれ順序対 $(\alpha^n, \mathcal{D}^n)$, (α^n, R) , (α^n, R) (ただし $\alpha^n = (a_{i_1}, \dots, a_{i_n})$
, $\mathcal{D}^n \triangleq \Theta_{i_1} \times \dots \times \Theta_{i_n}$, $R \subset \mathcal{D}^n$, $R \in \mathcal{D}^n$) で定義される。

ここに $(\Lambda, \dots, \Lambda)$ および $(\alpha^n, \{ (\Lambda, \dots, \Lambda) \})$ はそれぞれ null 組,

null 関係と呼ばれる。

<定義5> 順序対 $(\mathcal{A}^N, \mathcal{C}^N)$, (\mathcal{A}^N, G) , $(\mathcal{A}^N, \mathcal{Q})$ (ただし $\mathcal{A}^N = (a_{i_1}, \dots, a_{i_n})$, $\mathcal{C}^N \equiv \mathcal{D}_{i_1} \times \dots \times \mathcal{D}_{i_n}$, $G \subset \mathcal{C}^N$, $\mathcal{Q} \in \mathcal{C}^N$) のそれぞれを特に汎直積, 汎関係, 汎組と呼び, しばしば \mathcal{C}^N , G , \mathcal{Q} と略記する。

以上の諸定義により汎関係 G (generalized relation) における \mathcal{D} B 汎化の視野 (scope) として, 単一の \mathcal{A} の中で考察していることになる。

3. 関係演算系

3.1 タグ関係と条件付関係

G 上の関係代数系として, まず次のような関係演算を導入する。

<定義6> n 項組 $(\mathcal{A}^n, \mathcal{R})$, n 項関係 (\mathcal{A}^n, R) の属性識別子 \mathcal{A}^m による射影はそれぞれ次のように定義される:

$$(\mathcal{A}^m, \mathcal{R}[\mathcal{A}^m]) \triangleq \begin{cases} (\mathcal{A}^m, (\mathcal{R}_{i_1}, \dots, \mathcal{R}_{i_m})) & \text{ただし } 1 \leq m \leq n, \forall i (a_{i_j} \in \{a_{i_1}, \dots, a_{i_n}\}), \\ (\mathcal{A}^m, ()) & \text{さもなくば。} \end{cases}$$

$$(\mathcal{A}^m, R[\mathcal{A}^m]) \triangleq (\mathcal{A}^m, \{ \mathcal{R}[\mathcal{A}^m] \mid \mathcal{R} \in R \}).$$

<定義7> n 項関係 (\mathcal{A}^n, R) の条件値 (conditioning value) $\mathcal{D}^m \triangleq (d_{i_1}, \dots, d_{i_m})$ ($\forall i (i \in \{1, \dots, m\} \Rightarrow d_{i_j} \in \mathcal{D}_{i_j})$) によって条件付された部分関係 (sub-relation), 条件付関係 (conditioned relation), タグ関係 (tag relation) はそれぞれ次のように定義される:

$$(\mathcal{A}^n, R[\mathcal{A}^m \mid \mathcal{A}^m = \mathcal{D}^m]) \triangleq (\mathcal{A}^n, \{ \mathcal{R} \mid (\mathcal{R} \in R) \wedge (\mathcal{R}[\mathcal{A}^m = \mathcal{D}^m]) \}),$$

$$(\mathcal{A}^n - \mathcal{A}^m, R[\mathcal{A}^n - \mathcal{A}^m \mid \mathcal{A}^m = \mathcal{D}^m]) \triangleq (\mathcal{A}^n - \mathcal{A}^m, (R[\mathcal{A}^m \mid \mathcal{A}^m = \mathcal{D}^m])[\mathcal{A}^n - \mathcal{A}^m]),$$

$$(\mathcal{A}^m, R[\mathcal{A}^m = \mathcal{D}^m]) \triangleq (\mathcal{A}^m, (R[\mathcal{A}^m \mid \mathcal{A}^m = \mathcal{D}^m])[\mathcal{A}^m]).$$

しかし, 条件付関係とタグ関係の直積の直和が部分関係の直和となり, それらが元の関係 R を構成するという性質が成立する。

<性質1> ($\Sigma \oplus$: Hoare の意味の直和)

$$R[\mathcal{A}^n] = \sum_{\mathcal{D}^m \in \mathcal{C}^m} \oplus R[\mathcal{A}^n \mid \mathcal{A}^m = \mathcal{D}^m] = \sum_{\mathcal{D}^m \in \mathcal{C}^m} \oplus R[\mathcal{A}^n - \mathcal{A}^m \mid \mathcal{A}^m = \mathcal{D}^m] \times R[\mathcal{A}^m = \mathcal{D}^m].$$

3.2 θ -比較可能性による条件付演算

定義域間の演算の妥当性を検査するため θ -比較可能性という概念を導入し, それを用いた条件付関係演算を定義する。

<定義8> 必ずしも異なる定義域 $\mathcal{D}_{i_1}, \mathcal{D}_{i_2}$ ($\in \{\mathcal{D}_1, \dots, \mathcal{D}_n\}$) に対して, それらの基本型 $\Delta_{i_1}, \Delta_{i_2}$ ($\in \Delta$) が同一である場合のみ 1次元 θ -比較演算が, 次のように定義される。それ以外の場合, 演算は "未定義" である。

(i) Δ_{i_1} ($= \Delta_{i_2}$) 上において, θ として同値関係を許容する場合, 任意の d_{i_1} ($\in \mathcal{D}_{i_1}$), d_{i_2} ($\in \mathcal{D}_{i_2}$) に対して, " $d_{i_1} \theta d_{i_2}$ " は true あるいは false である。

(ii) Δ_{i_1} ($= \Delta_{i_2}$) 上において, θ として順序関係も許容する場合, 次のことが成立する。

(ii-1) 任意の d_{i_1} ($\in \mathcal{D}_{i_1} - \{\lambda\}$), d_{i_2} ($\in \mathcal{D}_{i_2} - \{\lambda\}$) に対して, " $d_{i_1} \theta d_{i_2}$ " は true あるいは false である。

(ii - ii) それ以外の場合, " $d_i \theta d_i$ " は "未定義" である.

全く同様に $\theta \triangleq (\theta_1, \dots, \theta_m)$ に対する m 次元 θ -比較演算 " $d_i^m \theta d_i^m$ " を, 次のように定義する:

$$d_i^m \theta d_i^m \triangleq (d_i \theta d_i) \wedge \dots \wedge (d_m \theta d_m).$$

ここでの "未定義" は演算結果の未定義なので \wedge と区別すべきであり, それに対するブール演算は別途, 目的に応じて定義されるべきである.

<定義9> n 項関係 (α^n, R) , 属性識別子 $\alpha_i^m, \alpha_i^m (c \alpha^n)$, 比較演算子 θ が与えられ, α_i^m, α_i^m が Δ に関し θ -比較可能ならば, θ 部分関係 $\cdot \theta$ 条件付関係 $\cdot \theta$ タグ関係はそれぞれ次のように定義される:

$$(\alpha^n, R [\alpha^n | \alpha_i^m \theta \alpha_i^m]) \triangleq (\alpha^n, \{ r \mid (r \in R) \wedge (r. \alpha_i^m \theta r. \alpha_i^m) \}),$$

$$(c \alpha^n - \alpha_i^m, R [c \alpha^n - \alpha_i^m | \alpha_i^m \theta \alpha_i^m]) \triangleq (c \alpha^n - \alpha_i^m, (R [\alpha^n | \alpha_i^m \theta \alpha_i^m]) [c \alpha^n - \alpha_i^m]),$$

$$(c \alpha_i^m, R [c \alpha_i^m | \alpha_i^m \theta \alpha_i^m]) \triangleq (c \alpha_i^m, (R [\alpha^n | \alpha_i^m \theta \alpha_i^m]) [c \alpha_i^m]).$$

3. 3 汎関係上の集合演算

<定義10> G 上の任意の関係 (α^m, R) , $(\alpha^{m'}, R')$ に対して, 和集合・共通集合・差集合という集合演算はそれぞれ次のように定義される:

$$(c \alpha^m, R \cup R') \triangleq (c \alpha^m, \{ g \mid (g \in G) \wedge ((g. \alpha_i^m \in R) \vee (g. \alpha_i^{m'} \in R')) \}),$$

$$(c \alpha^m, R \cap R') \triangleq (c \alpha^m, \{ g \mid (g \in G) \wedge (g. \alpha_i^m \in R) \wedge (g. \alpha_i^{m'} \in R') \}),$$

$$(c \alpha^m, R - R') \triangleq (c \alpha^m, \{ g \mid (g \in G) \wedge (g. \alpha_i^m \in R) \wedge (g. \alpha_i^{m'} \notin R') \}).$$

3. 4 汎関係上の導出関係

汎関係 G 上の関係代数の性質を明示するため <定義11> ~ <定義14> を準備する.

<定義11> $\mathcal{D}_{ij} (i \in \{1, \dots, n\}, j = 1 \dots n)$ を n 個の異なる定義域とする. それらの部分集合 $\mathcal{D}_{ij} (c \mathcal{D}_{ij})$ から成る集合 $\{d_{i_1}, \dots, d_{i_m}\}$ より, 必ずしも異なるない m 個の集合 $\mathcal{D}_{i_k} (k = 1 \dots m)$ を取出す. この時, 関係直積 $\prod_{k=1}^m (a_{i_k}, \mathcal{D}_{i_k})$ は次のように定義される:

(i) $\exists k (k \in \{1, \dots, m\}) \wedge (\mathcal{D}_{i_k} = \{\wedge\})$ ならば, 対応する定義域は全て縮約される.

(ii) $\exists k_1, \exists k_2 (k_1, k_2 \in \{1, \dots, m\}) \wedge (k_1 \neq k_2) \wedge (i_{k_1} = i_{k_2})$ ならば, 一意識別可能な属性を手がかりに対応する定義域は互いに単一化される.

(iii) それ以外の場合, 演算子 " \otimes " は演算子 " \times " に置換される.

<定義12> 関係直積 $\prod_{k=1}^m (a_{i_k}, \mathcal{D}_{i_k})$ を $\prod_{j=1}^n (a_{ij}, \mathcal{D}_{ij})$ 内に埋込まれた部分空間と呼び, $\prod_{k=1}^m (a_{i_k}, \mathcal{D}_{i_k}) \subseteq \prod_{j=1}^n (a_{ij}, \mathcal{D}_{ij})$ と表記する.

<定義13> 埋込み部分空間 $\prod_{k=1}^m (a_{i_k}, \mathcal{D}_{i_k})$ が l 項直積 $(\mathcal{B}^l, \mathcal{D}^l)$ (ただし $l \leq m$, $\mathcal{B}^l = (b_1, \dots, b_l)$, $\mathcal{D}^l = \mathcal{D}_1 \times \dots \times \mathcal{D}_l (i, \mathcal{D}_i \in \{\mathcal{D}_1, \dots, \mathcal{D}_n\})$) に還元され, しかも $R \subset \mathcal{D}^l$ とする. この時 (\mathcal{B}^l, R) を $\prod_{j=1}^n (a_{ij}, \mathcal{D}_{ij})$ 内に埋込まれた 部分関係 と呼び, $(\mathcal{B}^l, R) \subseteq \prod_{j=1}^n (a_{ij}, \mathcal{D}_{ij})$ と表記する.

<定義14> 汎関係 G 上の 導出関係 は, 次のように帰納的に定義される:

(i) 汎関係 $(c \alpha^m, G)$ は導出関係である.

(ii) n 項関係 (α^n, R) が導出関係ならば, その射影・部分関係・条件付関係・タグ関係・ θ 部分関係・ θ 条件付関係・ θ タグ関係も導出関係である.

(iii) n 項関係 (\mathcal{A}^n, R) , m 項関係 (\mathcal{A}^m, R') が導出関係なす, それらの和集合・共通部分・差集合も導出関係である。

以上の諸準備の下, G に対する検索処理系・更新処理系に対して, 次の二命題が成立することが証明された。

〈命題1〉 汎関係 G 上の任意の導出関係は, 関係直積 $(\mathcal{A}^n, \prod_{i=1}^n G[(a_i)])$ 内に埋込まれた部分関係となる。

〈命題2〉 汎関係 G が時間 t ($t \in T$) と共に変動したとしても, その上の任意の導出関係は, 関係直積 $(\mathcal{A}^n, \prod_{i=1}^n \cup_{t \in T} G_t[(a_i)])$ 内に埋込まれた部分関係である。

3.5 Codd の関係代数系との比較

R_1, \dots, R_L を Codd の意味の関係とし, それらの関係名を M_1, \dots, M_L とする。これらの関係を汎関係 G に埋込むため, 次の要請をする。

〈要請1〉 $\exists r ((r \in \{1, \dots, N\}) \wedge (\forall i (i \in \{1, \dots, L\} \Rightarrow M_i \in \mathcal{D}_r))) \wedge (\exists \Delta ((\Delta \in \{1, \dots, M\}) \wedge (\Delta \neq r) \wedge \neg (\forall i (i \in \{1, \dots, L\} \Rightarrow M_i \in \mathcal{D}_\Delta)))$ 。この要請は, ユニークな関係名定義域の存在の仮定である。しかるば, 次の性質が存在する。

〈性質2〉 任意の l ($l \in \{1, \dots, L\}$) に対して, $(\mathcal{D}_{M_l}, R_{M_l}) = (\mathcal{D}_G, \{g \cdot \mathcal{D}_{M_l} \mid (g \in G) \wedge (g \cdot (a_r) = M_l)\})$ 。

Codd の関係代数系の基本演算である射影・商・ θ 制約・ θ 結合と汎関係上の基本演算との対応関係を表示すると表1・表2のようになる。

	Our relational operation	Codd's relational operation
projection	$R[a]$	$R[a]$
conditioning	$R[a b=d]$	$(R[a]) [b=b] \{d\} \times \{d\}$
	$R[a-b b=d]$	$(R[a]) [b=b] \{d\}$
	$R[b=d]$	$\{d\}$
θ -conditioning	$R[a boc]$	$(R[boc]) [a]$
	$R[a-b boc]$	$(R[boc]) [a-b]$
	$R[boc]$	$(R[boc]) [b]$

表1. 汎関係上の関係演算のCoddの関係代数系による表現

	Codd's operation	Its representation by our notation
projection	$R_1[a_1]$	$\{r \cdot a_1 \mid (r \in G) \wedge (r \cdot (a_r) = N_1)\}$
division	$R_1[a_1 \div a_2] R_2$	$\{r \cdot (a_{N_1} - a_1) \mid (r \in G) \wedge (r \cdot (a_r) = N_1) \wedge (G[a_2 a_r = N_2] \subseteq g_{R_1}(r \cdot (a_{N_1} - a_1)))\}$ where $g_{R_1}(r \cdot (a_{N_1} - a_1)) = \{r \cdot a_1 \mid r \cdot a_{N_1} \in G[a_{N_1} a_r = N_1]\}$
θ -restriction	$R_1[a_1 \theta a_1']$	$\{r \cdot a_{N_1} \mid (r \in G) \wedge (r \cdot (a_r) = N_1) \wedge (r \cdot a_1 \theta r \cdot a_1')\}$
θ -join	$R_1[a_1 \theta a_2] R_2$	$\{r \cdot a_{N_1} \wedge r \cdot a_{N_2} \mid (r \in G) \wedge (r \cdot (a_r) = N_1) \wedge (r' \in G) \wedge (r' \cdot (a_r) = N_2) \wedge (r \cdot a_1 \theta r' \cdot a_2)\}$

表2. Codd の関係代数系の汎関係論理系表現

また次のような性質が存在する。

〈性質3〉 G 上の関係代数系は, θ 結合演算欠如のため関係完備でないが, 自然な結合演算に関しては関係完備である。

3. 6 応用ユーザに対する見方

Vehicle ::=

ID#	Manu.	Price	Weight	Medium category	Propulsion category
ID ₁	Mazda	65.4	10.5	land	motorized
ID ₂	Schwin	3.5	0.1	land	man powered
ID ₃	Boeing	7900.	840.	air	motorized
ID ₄	Aqua Co.	12.2	1.9	water	wind propelled
ID ₅	Gyro Inc.	650.	150.	air	motorized

図4. 乗物DB

Smith's のデータ抽象化と比較するため図5に示されるような乗物のカテゴリ-樹を考える。このVehicle DBの実現値が図4で示されているとする。この時、彼のgeneralization抽象化の見方でカテゴリ-motorizedを見ると、図6のような表が与え

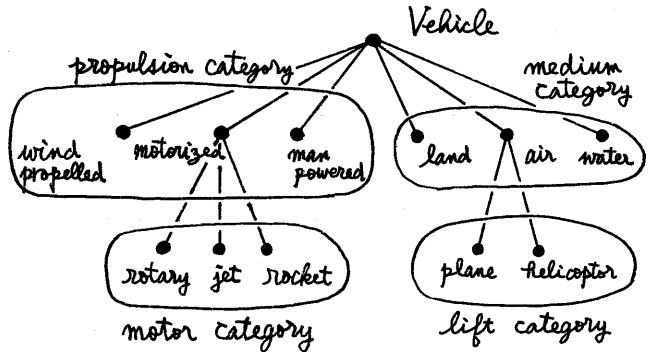


図5. 乗物カテゴリ-

Motorized vehicle ::=

ID#	Manu.	Price	Weight	Horse power	Fuel capacity	Motor category
ID ₁	Mazda	65.4	10.5	150	300	rotary
ID ₃	Boeing	7900.	840.	9600	2600	jet
ID ₅	Gyro Inc.	650.	150.	1500	2000	rotary

られる。すなわちカテゴリ-一化されたDBは、もとのDBのrefinementとなっている

図6. 原動機付乗物

ID#: iden. num.	M: manu- facturer	P: price	W: weight	HP: horse power	FC: fuel capacity	MA: max. alti- tude	TD: take- off dist.	LC: lift cate- gory	MFC: motor cate- gory	MC: medium cate- gory	PC: propul- sion cate- gory
		10000¥	kg	HP	l	m	m				
V1	Bell	11000	617	317	290	5200	0	heli- copter	turbine	air	motorized
V2	Tokyu Car	6000	28000	190	A	A	A	A	elec. motor	land	motorized
V3	Bridge- stone	4.6	20	A	A	A	A	A	A	land	man powered
V4	Toyota	95.5	1205	90	50	A	A	A	piston	land	motorized
V5	Yamaha	61	175	A	A	A	A	A	A	water	wind propelled
V6	Cessna	1500	580	150	160	4000	465	plane	piston	air	motorized
V7	Nippon Univ.	195	37	A	A	3	20	plane	A	air	man powered

図7. 汎関係Vehicle DB

他方、同様のVehicle DBを汎関係表現すると図7のようになる。この汎関係DBを、Pascal-like dddでデータ定義すると図8のようになる。ここにタグ関係・条件付関係はtag宣言とcase構文で示してあるが、

```

define relation as power set of
record ID#: identification number;
M: manufacturer;
P: price;
W: weight;
tag MC: medium category;
tag PC: propulsion category;
case MC of air:
record MA: maximum altitude;
TD: take-off distance;
tag LC: lift category
end record
end case;
case PC of motorized:
record HP: horse power;
FC: fuel capacity;
tag MFC: motor category
end record
end case
end record
    
```

図8. Vehicle DB に対するデータ ベース宣言例

Medium category, Propulsion categoryはそれぞれMC = air, PC = motorizedの場合しか例示してない。また図7の模式表現が図9であり、図5と比較すれば分かるように、汎関係表現は最も洗練(refinement)されたDBであり、かつ属性間のつながりが、階層的でなく平坦である。

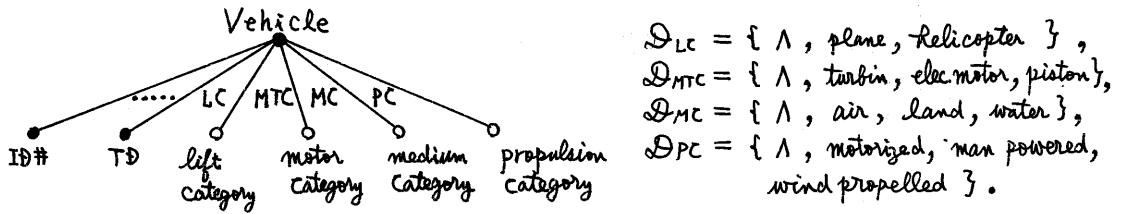


図9. 汎関係における属性の構造と各属性に対応する定義域

汎関係 VehicleDB に対するデータ定義(図8)においてタグの張られている属性MC (medium category) に注目する。属性MCの条件値 land, water による条件付関係が、図10の(a), (b)である。更にそれらから、図8のデータ宣言に従って、属性のフィルタリングを行なった結果が、図11である。

ID#: iden. num.	M: manufacturer	P: price	W: weight	HP: horse power	FC: fuel capacity %	MA: max. altitude m	TD: take-off dist. m	LC: lift category	MTC: motor category	PC: propulsion category
v2	Tokyu Car	6000	28000 kg	190 HP	A	A	A	A	elec. motor	motorized
v3	Bridgestone	4.6	20	A	A	A	A	A	A	man powered
v4	Toyota	95.5	1205	90	50	A	A	A	piston	motorized
v6	Yamaha	61	175	A	A	A	A	A	A	wind propelled

(a) Conditioned by MC=land

(b) Conditioned by MC=water

図10. medium category による条件付関係

ID#: iden. num.	M: manufacturer	P: price	W: weight	HP: horse power	FC: fuel capacity %	MTC: motor category	PC: propulsion category
v2	Tokyu Car	6000	28000 kg	190 HP	A	elec. motor	motorized
v3	Bridgestone	4.6	20	A	A	A	man powered
v4	Toyota	95.5	1205	90	50	piston	motorized
v6	Yamaha	61	175	A	A	A	wind propelled

(a) for MC=land

(b) for MC=water

図11. 条件付関係のddlによるフィルタリング

4. スキーマの進化

4.1 関係スキーマにおける従属関係の理論

RDBの関係スキーマに関する中心的課題は、完全性規約(integrity constraints)の一種であるFD・MVなどをア・プリオリに与え、これらと妥当な推論規則を用い、種々の正規化条件を満足するようにRDBを分解し設計する指針を与えることである。これにより正規化に伴う冗長性を防ぎ更新不整合が起らないような分解法を与えると言われている。しかしながら現在までに議論された従属関係の理論(5)~(7)においては、FD・MVの推論規則を別個のものとして扱っているのが、上林氏の指摘した場の問題が明示されておらず、種々の議論を巻起している。本節では、汎関係上でスキーマをどのように保存するか

ついで考察した結果， R が B の合成に伴う F ・ MVD の保存という立場から，属性の値の存在する場を明示する必要性を指摘し，新しく条件付関数従属（ CF ）・条件付多値従属（ $CMVD$ ）という概念を形式化し，従属関係の理論における“場の発生と消滅”の問題の部分的解決を試みるものである。

4.2 条件付従属関係の定式化

〈定義15〉 関係の族 \mathcal{R} が次のような条件を満足する時，与えられた場の条件（定義域） $\mathcal{D}_x, \mathcal{D}_y$ において属性識別子 X が属性識別子 Y に条件付関数従属するといひ， $(X \rightarrow Y / \mathcal{D}_x, \mathcal{D}_y)$ と表記する：

$$\forall R (R \in \mathcal{R} \Rightarrow \forall P, P' (P, P' \in R \Rightarrow \forall \alpha ((\alpha \in \mathcal{D}_x) \wedge (P.X = \alpha) \wedge (P'.X = \alpha) \Rightarrow \exists \beta ((\beta \in \mathcal{D}_y) \wedge (P.Y = \beta) \wedge (P'.Y = \beta))))).$$

〈定義16〉 関係の族 \mathcal{R} が次のような条件を満足する時，与えられた場の条件（定義域） $\mathcal{D}_x, \mathcal{D}_y, \mathcal{D}_z$ において属性識別子 X が属性識別子 Y と Z に条件付多値従属するといひ， $(X \rightarrow Y | Z / \mathcal{D}_x, \mathcal{D}_y, \mathcal{D}_z)$ と表記する：

$$\forall R (R \in \mathcal{R} \Rightarrow \forall P, P' (P, P' \in R \Rightarrow \forall \alpha ((\alpha \in \mathcal{D}_x) \wedge (P.X = \alpha) \wedge (P'.X = \alpha) \Rightarrow \exists P'', P''', \beta, \beta', \gamma, \gamma' ((P'', P''' \in R) \wedge (\beta, \beta' \in \mathcal{D}_y) \wedge (\gamma, \gamma' \in \mathcal{D}_z) \wedge (P''.X = P'''.X = \alpha) \wedge (P''.Y = P'.Y = \beta) \wedge (P'''.Y = P'.Y = \beta') \wedge (P''.Z = P'.Z = \gamma) \wedge (P'''.Z = P'.Z = \gamma')))).$$

さて二変数述語 $m(r, \xi)$ ($\triangleq r \in \xi$) と三変数述語 $e(r, \mathcal{V}, \mathcal{V})$ ($\triangleq r.V = \mathcal{V}$) を導入し，上記 CF ・ $CMVD$ の定義を置換すれば，次のようになる。

〈定義15'〉 $\forall R, P, P', \alpha \exists \beta (m(R, \mathcal{R}) \wedge m(P, R) \wedge m(P', R) \wedge m(\alpha, \mathcal{D}_x) \wedge e(P, X, \alpha) \wedge e(P', X, \alpha) \Rightarrow m(\beta, \mathcal{D}_y) \wedge e(P, Y, \beta) \wedge e(P', Y, \beta)).$

〈定義16'〉 $\forall R, P, P', \alpha \exists P'', P''', \beta, \beta', \gamma, \gamma' (m(R, \mathcal{R}) \wedge m(P, R) \wedge m(P', R) \wedge m(\alpha, \mathcal{D}_x) \wedge e(P, X, \alpha) \wedge e(P', X, \alpha) \Rightarrow m(P'', R) \wedge m(P''', R) \wedge m(\beta, \mathcal{D}_y) \wedge m(\beta', \mathcal{D}_y) \wedge m(\gamma, \mathcal{D}_z) \wedge m(\gamma', \mathcal{D}_z) \wedge e(P, Y, \beta) \wedge e(P, Z, \gamma) \wedge e(P', Y, \beta') \wedge e(P', Z, \gamma') \wedge e(P'', X, \alpha) \wedge e(P'', Y, \beta) \wedge e(P'', Z, \gamma') \wedge e(P''', X, \alpha) \wedge e(P''', Y, \beta') \wedge e(P''', Z, \gamma)).$

4.3 条件付従属関係の性質とその汎関係上での保存

場の条件 $\mathcal{D}_x, \mathcal{D}_y, \mathcal{D}_z$ は関係の族 \mathcal{R} と独立に与えうる。そこで \mathcal{R} を先験的に与え， \mathcal{R} に依存する固定された場をそれぞれ次のようにして構成する：

$$\mathcal{D}_x^* \triangleq \bigcup_{R \in \mathcal{R}} R[X], \quad \mathcal{D}_y^* \triangleq \bigcup_{R \in \mathcal{R}} R[Y], \quad \mathcal{D}_z^* \triangleq \bigcup_{R \in \mathcal{R}} R[Z].$$

また族 \mathcal{R} の族 \mathcal{R} に対する固定された場も考えうる。

$$\mathcal{D}_x^{**} \triangleq \bigcup_{R \in \mathcal{R}} \mathcal{D}_x^*, \quad \mathcal{D}_y^{**} \triangleq \bigcup_{R \in \mathcal{R}} \mathcal{D}_y^*, \quad \mathcal{D}_z^{**} \triangleq \bigcup_{R \in \mathcal{R}} \mathcal{D}_z^*.$$

前者を固定された局所場 (local context)，後者を固定された大域場 (global context) と呼ぶ。しからば固定された局所場 $\mathcal{D}_x^*, \mathcal{D}_y^*, \mathcal{D}_z^*$ で成立する CF ・ $CMVD$ が，従来の F ・ MVD の定義と一致する。そこで前節の冠頭標準形による CF ・ $CMVD$ の定義を用い，述語論理の諸定理および竹島の resultant 定理^{33), 34)}等を用いれば，以下の諸定理・諸命題が証明できる³⁵⁾。

< 定理 1 > $\subset F \supset D$ は、固定された局所場において Armstrong の 6 公理 9) を満足する。

< 定理 2 > $\subset M \supset V \supset D$ は、固定された局所場において Fagin の 3 公理 13) を満足する。

< 定理 3 > 与えられた $(X \stackrel{R}{\rightarrow} Y / \mathcal{D}_x, \mathcal{D}_y)$, $(X' \stackrel{R'}{\rightarrow} Y' / \mathcal{D}_x', \mathcal{D}_y')$ に対して $(X \stackrel{R \cup R'}{\rightarrow} Y / \mathcal{D}_x \cup \mathcal{D}_x', \mathcal{D}_y \cup \mathcal{D}_y')$ とする必要十分条件は、 $(X \stackrel{R \cup R'}{\rightarrow} Y / \mathcal{D}_x \cap \mathcal{D}_x', \mathcal{D}_y \cap \mathcal{D}_y')$ が成立することである。

汎関係 G 上では、< 要請 1 > で仮定されたユニークな関係名定義域 \mathcal{D}_r に関する次の二命題が成立する。

< 命題 3 > $(X \stackrel{R}{\rightarrow} Y / \mathcal{D}_x, \mathcal{D}_y)$ が成立すれば、汎関係 G の族 \mathcal{G} 上では、 $(\{a_r\} \cup X \stackrel{R}{\rightarrow} Y / \mathcal{D}_r \times \mathcal{D}_x, \mathcal{D}_y)$ が成立する。

< 命題 4 > $(X \stackrel{R}{\rightarrow} Y \mid Z / \mathcal{D}_x, \mathcal{D}_y, \mathcal{D}_z)$ が成立すれば、汎関係 G の族 \mathcal{G} 上では、 $(\{a_r\} \cup X \stackrel{R}{\rightarrow} Y \mid Z / \mathcal{D}_r \times \mathcal{D}_x, \mathcal{D}_y, \mathcal{D}_z)$ が成立する。

上記命題より、汎関係 G 上で成立する $\subset F \supset D \cdot \subset M \supset V \supset D$ の構成法が与えられる。固定された大域場で成立する $\subset F \supset D \cdot \subset M \supset V \supset D$ の理論、および固定されなりの場の理論を構築していくことが、今後の研究課題といえる。

5. おわりに

汎関係 $D \supset B$ の長短を、data syntax · data semantics · data pragmatics という見地から論じてみる。

データ統語論的には汎関係 $D \supset B$ は、情報代数と同じく唯一の情報空間を考慮しており、単一の直積型データ構造の内部に、tag(s) という条件の場合分けにより創成される条件付関係を、直和型データ構造として埋込んだものであり、Smith の考え方の直積と直和の考え方が逆転している。しかも汎関係においては、ddl でのタグ宣言により特定の視点でのクラスタリングが、自然な形で記述できる。何故ならタグによつて条件付された部分関係・条件付関係が、特定の分類基準で見た時のクラスタを表わしているからである。また Codd の関係 (連) のもっている潜在的情報 (たとえば関係名、属性集合、5W1H 等の条件に関する情報) を顕在化し、一枚の表形式データとして見せるので、概念スキーマ・外部スキーマ・内部スキーマのどのレベルのコーザにとつても、情報の表現形式としての自己記述性がある。

データ意味論的には「組 R が関係 R に属しますか？」という質問だけでなく、「組 R はどんな関係にありますか？」という「関係の意味」を問う質問が、汎関係では可能である。これは関係の実体でなく「名前」を値とする定義域が存在するからである。また「システム性能が 250 以上のハードウェアを提供している計算機会社はどこですか？」(例 2 を参照) とし、た質問にも答え得る。これは、ある属性方向への射影値が等しいものは同一の entity とみなす「entity 同一視の原理」を用いるか、あるいは $F \supset D$ (または Key) による値出し演算を用いればよい。いずれにせよ汎関係は Question-Answering 志向があり、QA に対する推論 (意味論的制約) にスキーマ情報が利用できる。

最後にデータ・プラグマティック的には、汎関係の実働化イメージを堅固に出さんとするなら、< 属性、量 > 表現の立場を更に徹底し、< 定義 11 > に見られるような情報 (データ) 圧縮の問題をも、と考察しなければならぬ。その

際、一様なアクセスを保障するような key の取扱いが困難であろう。またデータの動的取扱い時、すなわち更新処理時においては、属性の追加・削除モードにおける dml の負担が、かなり大きくなりそうなので、何らかの対策が必要である。

以上の諸議論を要約するに、汎関係 DB は人間の知識の集大成である DB に対して一様かつ平坦な見え方を与え、しかも方法の集大成である、ある分野固有の視座にそって切断する条件付演算という切断法を提供する。今後の研究課題としては、条件付従属を用いる汎関係の分解法を与え、より実働化できる形に接近することで、およびア・プリアリなスキーマ利用法ばかりでなく、ア・ポステリオリなスキーマ利用法（例えば、F₀による推論とか F₀の発見）等についての研究が重要である。

〔謝辞〕 本研究を推進するに当り、熱心に討論していただき、多くの貴重な助言をしてくださいました竹島卓（富士通国際研）、高橋浩（富士通のIPSG部）両氏に感謝の意を表します。また懇切丁寧にご指導していただきました北川敏男所長（富士通国際研）に謝意を表します。更に本研究に対して適切なコメントをいただいた林達也室長（富士通研）、杉山公造博士（富士通国際研）、上林弥彦先生（京大）、總鷹良介博士（SSU）、増永良文先生（東北大）、田中護先生（北大）、有沢博先生（横浜国大）および同僚の国際研諸氏に感謝致します。

〔文献〕

- 1) Codd, E.F.: Relational completeness of data base sublanguages, *Data Base Systems*, 1971.
- 2) Liskov, B. & Zilles, S.: Programming with abstract data types, *Proc. of Symp. on VLSL*, 1974.
- 3) Smith, J. M. & Smith, D. C. P.: Database Abstractions: Aggregation and Generalization, *TODS*, Vol. 2, 1977, 105-133.
- 4) Hoare, C. A. R.: Notes on data structuring, in *Structured Programming*, Academic Press, 1972.
- 5) Schmid, H. A. & Swenson, J. R.: On the Semantics of the Relational Data Model, *ACM-SIGMOD*, 1975, 211-223.
- 6) Codd, E.F.: Further normalization of the data base relational model, *Data Base Systems*, 1971.
- 7) Delobel, C. & Casey, R. G.: Decomposition of a data base and the theory of Boolean switching functions, *IBM J. Res. Develop.*, 17-5, 1972, 374-386.
- 8) Bernstein, P. A.: Synthesizing Third Normal Form Relations from Functional Dependencies, *TODS*, 14, 1976, 277-298.
- 9) Armstrong, W. W.: Dependency structures of data base relationships, *Inf. Processing* 74, 1974, 580-583.
- 10) Kamayashi, Y.: Synthetic Approach to Relational Data Base Design Using Boolean Functions, *VLDB*, 1977.
- 11) Rissanen, J.: Independent Components of Relations, *TODS*, Vol. 2, 1977, 317-325.
- 12) 増永良文, 野口正一: 関係間関数依存性を考慮した関係データベースの構成法, *信学技報 AL* 77-36, 1977, 93-99.
- 13) Fagin, R.: Multivalued Dependencies and a New Normal Form for Relational Databases, *TODS*, Vol. 2, 1977, 262-278.
- 14) Beeri, C., Fagin, R. & Howard, J. H.: A complete axiomatization for functional and multivalued dependencies, *IBM Res. Rep.* RJ 1977.

- 15) Fagin, R.: *Dependency in a Relational Database and Propositional Logic*, IBM Res. Rep. RJ 1776, 1976.
- 16) 上林 弥彦他: リレーショナル・データベースにおけるいくつかの問題, DBMS クラス研究会, 東大, 1977.
- 17) Makinouchi, A.: *A Consideration on Normal Form of Not-necessarily normalized Relation in the Relational Data Model*, 3VLDB, 1977.
- 18) CODASYL: *An Information Algebra*, CACM 5, 1962.
- 19) Kobayashi, I.: *Information and Information Processing Structure*, Information System 1, 1975.
- 20) 橋 正明: プログラム管理およびユーザ言語管理機能を補強した DBMS-DPLS, 情報処理, 17-10, 1976.
- 21) Hatake, R. & Tsubaki, M.: *Self-Descriptive Relational Data Base*, 3VLDB, 1977.
- 22) 弘原海清: 単板・関係・網・階層型を含む複合論理データベースについて — 学際研究用 DBMS: GED DAS を例として —, 情報処理学会, データベース管理システム 5-1, 1978.
- 23) 小島 順: 線型代数, 日本放送出版協会, 1976.
- 24) 島内 剛一: プログラム言語論, 共立出版, 1972.
- 25) 植村 俊亮: データベースの関係モデル第3正規形批判, 情報処理, 18-1, 1977.
- 26) 總鷹 良介: リレーショナル・データベースにおける第3正規形について, 情報処理, 18-11, 1977.
- 27) 總鷹 良介, 波谷 政昭: データベースの関係図式, 情報処理, 17-10, 1976.
- 28) 有澤 博: リレーショナル・データベース上の領域間の決定関係についての考察, 情報処理, 17-11, 1976.
- 29) 小林 要: 汎関係データベース・モデル (1), 情報処理学会第18回全国大会, 1977.
- 30) 国藤 進: 汎関係データベース・モデル (2), 情報処理学会第18回全国大会, 1977.
- 31) 杉山 公造: 汎関係データベース・モデル (3), 情報処理学会第18回全国大会, 1977.
- 32) 北川 敏男: 汎関係圏の構造, 日本数学会, Oct. 10, 1977.
- 33) 竹島 卓: リレーショナル・データベースにおける関数従属関係と第3正規形化問題, 情報処理学会第18回全国大会, 1977.
- 34) 竹島 卓: 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, 情報処理学会, データベース管理システム研究会, 6-2, 1978.
- 35) 国藤 進: 条件付き従属を用いる関係 DB における従属関係の理論の一般化の試み, 昭和53年度電子通信学会総合全国大会, 1978.