

関係データベースシステムにおける 従属性を利用したデータの表現について

上林 弥彦^{*} 田中 克己⁺ 武田 浩一^{*} 矢島 脩三^{*}

^{*} 京都大学 工学部

⁺ 神戸大学 教養部

1. まえがき 関係データベースシステムにおける関係の正規化は、⁽⁶⁾単純であるが強力な利用者言語の提供や、更新による不整合の低減等を実現するうえで重要な概念である。しかし正規形によるデータの表示は利用者にとってわかり易いものとは言えず、FORMを用いたデータの表示等の非正規表現の研究が行なわれている。^{(11),(12)}即ち、利用者インタフェースとして正規化された関係と非正規表現の変換を考えることが必要となる。非正規表現はデータ自体とそのデータが満たす意味的制約をコンパクトに表現していると考えられ、言い換えれば、与えられたデータが満たす意味的制約をできるだけ反映した非正規表現を見つけることが、わかり易いデータの表示において重要である。

本稿では意味的制約として数学的な扱いが容易なデータ従属性を考え、非正規表現として GROUP-BY 操作⁽⁵⁾によって得ら

れる水平方向の階層表現を用いた場合の両者の関連について考察している。本稿での考察結果は、以下のように関係データベースシステムに適用できる。

- (i) データベースシステムでは、データは正規形の関係と、その関係上で成立する従属性集合によって表現される。
- (ii) 利用者がデータを出力するときには、そのデータが満たす従属性を反映した非正規表現を用いて表示できる。

2節では基本的事項を示し、3節では関数従属性と水平方向の階層表現、4節では結合従属性と水平方向の階層表現について考察している。

2. 基本的事項 関係 R は属性集合 $\underline{R}$ の定義域値の組み合わせによって得られる組の有限集合とする。 $\underline{R}$ は R に対応する関係スキームを表わす。アルファベットの初めの文字 $A, B, \dots$ で1つの属性を示し、 $\dots, X, Y, Z$ で属性集合を示す。属性あるいは属性集合の和はその連接によって示す。関係 R 上の属性集合 X と組 α について、 α の X -値とは、組 α の X に対応する部分の値とする。関係 R の $X \subseteq \underline{R}$ 上の射影を $R[X]$ で示し、 R から X 以外の属性をすべて除いた関係とする。2つの関係 R_1, R_2 の自然結合を $R_1 * R_2$ で示し、 $\underline{R_1} \cup \underline{R_2}$ 上の関係とする。 $R_1 * R_2$ は、 $\alpha[\underline{R_1}] \in R_1$ かつ $\alpha[\underline{R_2}] \in R_2$ のとき、そのときのみ組 α を含む。ただし $\alpha[\underline{R_i}]$ は α の $\underline{R_i}$ -値を示す。

データベーススキームは関係スキームの集合であり，通常は意味的制約の集合 C をもつ。意味的制約のうちで以下の従属性と呼ばれるクラスが重要である。関数従属性 (FD) $X \rightarrow Y$ は，各 X -値に対して， Y -値がその X -値によって一意に定められるとき，またそのときのみ成立する。結合従属性 (JD) $*[Y_1, Y_2, \dots, Y_n]$ は， $R_1 = \bigcup_{i=1}^n Y_i$ なる関係 R_1 において， R_1 が常に $Y_1, Y_2, \dots, Y_n$ に分解でき， $R_1 = \bigstar_{i=1}^n R_1[Y_i]$ となるとき，そのときのみ成立する。与えられた従属性の集合 D について， D から論理的に導かれるすべての従属性の集合を D の閉包といい D^+ で表わす。

関係は正規形，非正規形のいずれでも表現されうる。前者では単純値のみが定義域値となり，後者では集合値や関係そのものが定義域値となつてよい。

3. 関数従属性を用いた水平階層表現 本節では関数従属性 (FD) の集合が与えられた場合に水平方向の階層表現を得る方法について考察する。水平方向の階層表現は GROUP-BY 操作によって得られ，このとき次の定理が成立する。

定理 1 関係 R の属性集合 X に対する GROUP-BY 操作によって得られる各部分関係において FD $\phi \rightarrow X$ が成立する。

各部分関係において X の値は定数であるから証明は明らかである。このとき R において成立する FD は各部分関係で成

立し，さらに $\phi \rightarrow X$ を加えた FD 集合から導かれる FD も各部分関係において成立する。

定理 2 もし関係 R において FD の鎖 $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n$ が成立するなら，この FD 鎖を反映した水平方向の階層表現が次の GROUP-BY 操作の系列によって得られる。

GROUP BY X_n ; GROUP BY $X_{n-1} - X_n$; ... ;

GROUP BY $X_2 - X_n - X_{n-1} - \dots - X_3$;

各 $X_i \rightarrow X_{i+1}$ に対して， X_{i+1} による GROUP-BY 操作は関係 R を X_i の値が互いに重複しないように分割できるので，各 GROUP-BY 操作により階層表現が得られることは明らかである。また， $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n$ が成立するとき， X_n による GROUP-BY 操作の結果，各部分関係で $\phi \rightarrow X_n$ が成立するため， $\{X_1 - X_n\} \rightarrow \{X_2 - X_n\} \rightarrow \dots \rightarrow \{X_{n-1} - X_n\}$ が成立することがわかる。これを繰り返せば定理 2 が得られる。

定理 2 を用いて得られる階層表現に対して，属性の位置を移動することで関連した属性集合をまとめることができる。このような属性間の階層を垂直方向の階層と呼ぶ。図 1 にその例を示す。図 1 (a) の関係 R で $AB \rightarrow CD \rightarrow EF$ という FD 鎖が成立し， $Y = DEB$ という関連属性集合が与えられたとき，図 1 (b) のような水平方向の階層表現を得たあとで，図 1 (c) のように属性を移動し，さらに属性 E の列を分割して，

図1 (d) のような結果を得ることができる。

次に、1つの関係に2つ以上のFD鎖が存在するときには、複数の関係によってFD鎖をそれぞれ反映させるが、FD鎖をまとめる必要があり、⁽¹⁰⁾ 後者の場合には下の定理を用いることができる。

定理3 もし関係 R において2つのFD鎖 $U_1 \rightarrow U_2 \rightarrow \dots \rightarrow U_n$ および $V_1 \rightarrow V_2 \rightarrow \dots \rightarrow V_m$ が成立するなら、次のFD鎖もやはり R において成立する； $W_1 \rightarrow W_2 \rightarrow \dots \rightarrow W_p$ ここで $p = \max(m, n)$ かつ $W_i = U_i V_i$ である。（もし $m > n$ なら $n < i \leq m$ に対して、 $W_i = V_i$ ， $m < n$ なら $m < i \leq n$ に対して $W_i = U_i$ である。）

以上の結果より、関係 R と R で成立するFD集合および関連属性集合が与えられた場合に水平方向および垂直方向の階層表現を生成する手続きが得られる。

Procedure 1 入力：関係 R ， R 上で成立するFD集合および関連属性集合 出力： R の水平および垂直方向の階層表現
 (1) FD集合からFD鎖の集合を求める。(2) FD鎖を変形し、関連属性集合がいずれかのFD鎖に含まれるようにする。
 (3) 各FD鎖に対して階層表現を生成し、関連属性集合をまとめる。

FD鎖が複数個あるときには、生成される階層表現が情報無損失であることを保証するために、汎関係のキー属性から

成る長さ 1 の FD 鎖が含まれることがある。^{(4),(9)} FD 鎖の最適化(くり返して現われる属性数が少なく, より長い FD 鎖を求めること)は FD に基づく関係スキームの合成手法に関連しており,^{(3),(8),(9)} ここでは省略する。

4. 結合従属性と水平階層表現

本節では結合従属性(JD)と水平階層表現の関連について考察する。JD が成立するときには, 属性値が直線的に対応する水平方向の階層表現を考えると都合がよく, 一般に直線的な対応を利用してコンパクトな表現が得られるが, 巡回的 JD⁽¹⁾ と呼ばれる JD に対しては効果が得られないことを示す。

補題 4 関係 R が $JD * [S_1, \dots, S_m]$ (ただし $R = \bigcup_{i=1}^m S_i$) を満たし, X が R の部分集合であるとする。このとき X による GROUP-BY 操作で得られる各部分関係においても $JD * [S_1, \dots, S_m]$ が成立し, さらに GROUP-BY 操作により, $FD \phi \rightarrow X$ が各部分関係で成立するため, X 以外の部分において $*[S_1 - X, \dots, S_m - X]$ が成立する。

証明は各部分関係で X が定数であることや, 文献(2)の結果を用いて容易に行なえる。補題 4 を用いて, GROUP-BY 操作と FD, JD の干渉問題を次のように示すことができる。

定理 5 関係 R が FD 集合 F と JD 集合 J を満足するものとする。このとき X による GROUP-BY 操作を施して得られる

各部分関係において，FD 集合 F' と JD 集合 J' が成立し，

$$F' = \{\emptyset \rightarrow X\} \cup \{Y \rightarrow Z \mid XY \rightarrow Z \in F^+\}$$

$$J' = \{*[S_1 - X, \dots, S_m - X] \mid *[S_1, \dots, S_m] \in J\}$$

である。

JD が与えられた場合に，どのような順で GROUP-BY 操作を行なえばよいかは，与えられた JD のハイパーグラフによる表現から定めることができる。例を図 2 に示す。今，関係 R を $R = ABCDEFG$ 上の関係とし， R で成立する JD を $*[ABC, CDE, AEF, AG]$ とする。JD に対応するハイパーグラフは図 2 (a) のようになる。もし属性 A について GROUP-BY 操作を施すと，属性 A 以外の部分では $*[BC, CDE, EF, G]$ が成立する。(図 2 (b)) もし最初に BC について GROUP-BY 操作を施すと， $*[DE, AEF, AG]$ が成立する。(図 2 (c)) A に対する GROUP-BY 操作ではハイパーグラフが 2 つの連結成分に分けられるのに対し， BC の場合には 1 つの連結成分のみであり，前者の方が直積的な対応による効果が大い。従って，ハイパーグラフ上でなるべく多くの連結成分にわかれるような GROUP-BY 操作から順に施せばよいことがわかる。図 3 (a) のような関係 R においては，まず A により GROUP-BY 操作を行ない，次に C により GROUP-BY 操作を行なう。それぞれの結果が図 3 (b), (c) に示される。ここで注意すべきことは，いずれの

GROUP-BY操作の場合も2つの連結成分を生じるが、CによるGROUP-BY操作はあまり効果的でない。これは、もともとABC, CDE, AEFという巡回的なJDの成分があるためである。以上の考察より、与えられたJDから直積的な属性値の対応を利用した水平方向の階層表現を生成する手続きが得られる。

Procedure 2 入力: 関係R, R上で成立するJD $j: * [S_1, S_2, \dots, S_m]$ 。ここで次の記法を用いる。H(j): j を表現するハイパーグラフ, $DEL(j, X) = * [S_1 - X, S_2 - X, \dots, S_m - X]$ 。
 (1) JDの巡回成分は1つにまとめて、その結果得られる非巡回的なJDを j とする。(2) H(j)からXを除いた結果、連結成分の数を増やすような最小のXを見つける。(3) XによるGROUP-BY操作をRに施す。(4) j を $DEL(j, X)$ におきかえる。(2), (3)をXが見つけられる間くりかえす。

5. 結論 本稿では意味的制約として関数従属性, 結合従属性というデータ従属性が与えられた場合に, GROUP-BY操作によって水平方向の階層表現, 属性値の直積的な対応を用いたコンパクトな水平方向の階層表現をそれぞれ生成できることを示した。関数従属性に基づいて生成される水平方向の階層表現には, 属性の位置を移動することにより関連属性集合をまとめるという垂直方向の階層表現を加えることができるが, 直積的な属性値の対応を用いた水平方向の階層表現では

属性を簡単に移動できない場合があり，結合従属性と関連属性集合を同時に反映した表現については今後の課題である。

一般に，実世界の結合従属性はほとんど非巡回的である⁽¹⁾という仮説に従えば，与えられた結合従属性を反映した水平方向の階層表現は4節の手続きによって生成できる。

謝辞 日頃御討論頂く矢島研究室の諸氏に感謝する。

参考文献

1. Beeri, C., Fagin, R., Maier, D., Mendelzon, A.O., Ullman, J.D. and Yannakakis, M. "Properties of Acyclic Database Schemes", Proc. 13th ACM STOC, pp.355-362, May 1981
2. Beeri, C. and Rissanen, J. "Faithful Representations of Relational Database schemes", IBM Research Report, RJ2772, January 1980
3. Bernstein, P.A. "Synthesizing Third Normal Form Relations from Functional Dependencies", ACMTODS, Vol.1, No.4, pp.277-298, December 1979
4. Biskup, J., Dayal, U. and Bernstein, P.A. "Synthesizing Independent Database Schemes", Proc. ACM-SIGMOD, pp.143-151, May-June 1979
5. Chamberlin, D.D. et al. "SEQUEL2: A Unified Approach to Data Definition, Manipulation and Control", IBM Technical Report, RJ1798, June 1976
6. Codd, E.F. "A Relational Model of Data for Large Shared Data Banks", CACM, Vol.13, No.6, pp.377-387, June 1970
7. Fagin, R., Mendelzon, A.O. and Ullman, J.D. "A Simplified Universal Relation Assumption and its Properties", IBM Technical Report, RJ2900, Nov. 1981
8. Kambayashi, Y. "Equivalence Key Problem of the Relational Data Base Model", Lecture Notes in Computer Science 75, pp.165-192, August 1978
9. Kambayashi, Y. "A New Synthetic Approach for Relational Database Design", Kyoto Univ. Dept. of Information Science, Yajima Lab. Research Report, (presented at AFIPS 1979 NCC), ER78-02, November 1978

10. Kambayashi, Y., Tanaka, K., Takeda, K. and Yajima, S. "Representation of Relations for Database Output Utilizing Data Dependencies", Proc. 15th Hawaii ICSS, pp.69-78, January 1982
11. Kitagawa, K., Kunii, T.L. and Ishii, Y. "Design and Implementation of a Form Management System APAD using ADABAS/INQ DBMS", Proc. COMPSAC, pp.324-334, November 1981
12. Luo, Y. and Yao, S.B. "Form Operation By Example - a Language for Office Information Processing", Proc. ACM-SIGMOD, pp.212-223, April 1981

A	B	C	D	E	F
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—

(a)

E	F	C	D	A	B
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—

(b)

F	C	Y			A
		D	E	B	
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—

(c)

F	C	Y			A
		D	E	B	
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—
—	—	—	—	—	—

(d)

図1

FD 鎖と関連属性集合の利用

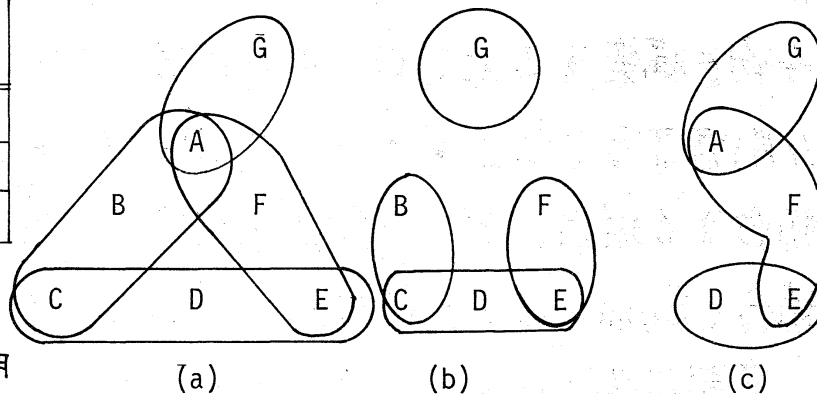


図2 JDのハイパーグラフによる表現

A	B	C	D	E	F	G
a	a	a	b	b	b	a
a	a	a	b	b	b	b
a	a	a	a	a	a	a
b	b	a	a	a	b	b
a	b	b	b	a	a	a
a	b	b	b	a	a	b

(a)

A	G	B	C	D	E	F
a	a	a	a	b	b	b
	b	b	b	b	a	a
		a	a	a	a	a
b	b	b	a	a	a	b

(b)

A	G	C	B	D	E	F
a	a	a	a	a	a	a
	b	b	b	b	a	a
b	b	a	b	a	a	b

(c)

図3 JDを利用した GROUP-BY 操作の適用