数理解析研究所講究録
第 458 巻 1982 年 172-181

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, …で1つの属性を示し，…, 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tについて，tのX－値とは，組t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，$t[\underline{R_1}] \in R_1$かつ$t[\underline{R_2}] \in R_2$のとき，そのときのみ組$t$を含む。ただし$t[\underline{R_i}]$は$t$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）$*[Y_1, Y_2, \cdots, Y_n]$ は，$\underline{R_1} = \bigcup_{i=1}^{n} Y_i$ なる関係 R_1 において，R_1 が常に $Y_1, Y_2, \cdots, Y_n$ に分解でき，$R_1 = \mathop{*}_{i=1}^{n} R_1[Y_i]$ 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ってよい。

<u>3．関数従属性を用いた水平階層表現</u>　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。

<u>定理１</u>　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$X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$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 $X_2 - X_n - X_{n-1} - \cdots - X_3$ ；

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，$X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n$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，$\{X_1 - X_n\} \rightarrow \{X_2 - X_n\} \rightarrow \cdots \rightarrow \{X_{n-1} - X_n\}$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か，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 $U_1 \to U_2 \to \cdots \to U_n$ および $V_1 \to V_2 \to \cdots \to V_m$ が成立するなら，次のFD鎖もやはりRにおいて成立する； $W_1 \to W_2 \to \cdots \to W_p$　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，属性値が直積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，一般に直積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 $*[S_1, \cdots, S_m]$ (ただし $\underline{R} = \bigcup_{i=1}^{m} S_i$)を満たし，Xが$\underline{R}$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 $*[S_1, \ldots, S_m]$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$*[S_1-X, \cdots, S_m-X]$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，

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

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

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$\underline{R}$ = ABCDEFG 上の関係とし，Rで成立するJDを*[ABC, CDF, 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きい。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：$*[S_1, S_2, \cdots, S_m]$。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，$DEL(j, X) = *[S_1-X, S_2-X, \cdots, S_m-X]$。(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

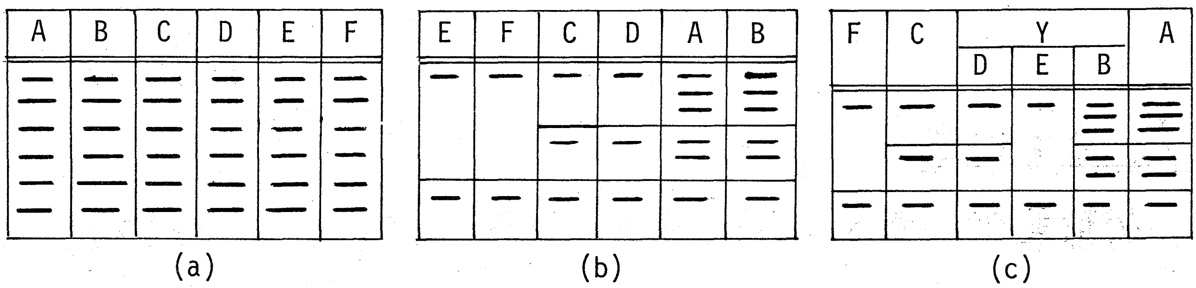
属性を簡単に移動できない場合があり，結合従属性と関連属性集合を同時に反映した表現については今後の課題である。一般に，実世界の結合従属性はほとんど非巡回的である(1)という仮説に従えば，与えられた結合従属性を反映した水平方向の階層表現は4節の手続きによって生成できる。

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



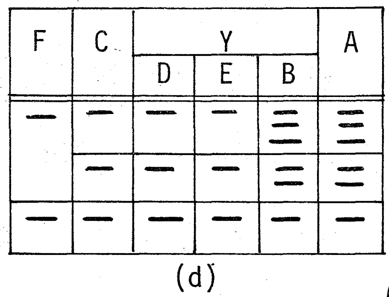


図1
FD鎖と関連属性集合の利用

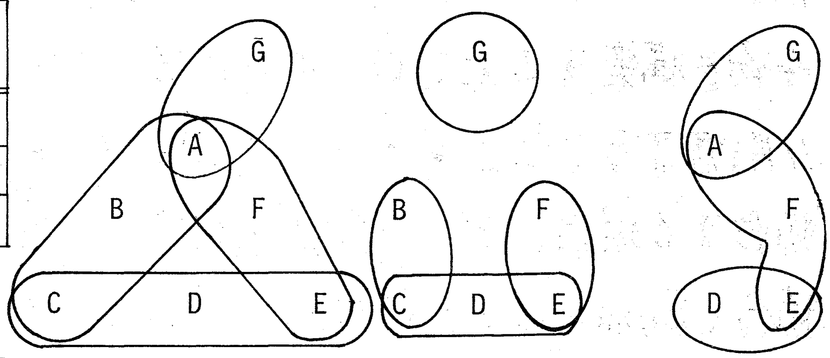


図2　JDのハイパーグラフによる表現

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 b	a b	a b
	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