

横切り分解によるデータ従属性の保持

京都大学工学部 上林弥彦 Yahiko Kambayashi

神戸大学教養部 田中克己 Katsumi Tanaka

1. まえかき

データベース設計において残された一つの重要な問題は、従属性集合の持つ閉路の扱いである。この問題に対しては従来よく用いられている射影分解では扱うことが困難であるため別の方法を検討する必要がある。関数従属性に対しては Sciore が属性追加による方法を発表している。⁽¹⁾ 本稿ではデータ表現の FORM の問題とも関係する⁽²⁾ 関係の横切りによる従属性集合の持つ閉路の取り扱いについて考察する。従属性の持つ閉路は算術処理の問題とも大きく関係しており重要である。⁽³⁾

2. 基本的定義

$A, B, \dots$ は一つの属性を表わし, $\dots, X, Y, Z$ は属性集合を表わすものとする。関係 R_i は属性集合 R_i に対して定義された組の有限集合である。 R_i を R_i に対する関係スキーマと呼ぶ。簡単のため、和集合を連接で示すものとする。たとえば、 $AB = \{A, B\}$, $XY = X \cup Y$ である。

組 t のうち属性集合 X に対応する部分のみをとる射影を、 $t[X]$ で表わす。関係の射影、選択、自然結合は次のように定義される。

$$\text{射影: } R[X] = \{t[X] \mid t \in R\}$$

$$\text{選択: } R[X = c] = \{t \mid t[X] = c, t \in R\}$$

$$\text{自然結合: } R * S = \{tu \mid t[R \cap S] = u[R \cap S], \\ t \in R, u \in S\}$$

関数従属性 (FD) : $X \rightarrow Y$ が関係 R で成立するといわれるのは、すべての組 t , s の組合せにおいて、 $t[X] = s[X]$ ならば $t[Y] = s[Y]$ となる場合である。FD のうち Y が一属性のものを基本 FD という。全関数従属性 (FFD) は、 $X \rightarrow Y$ と表わされ、 $\forall A \in Y \Rightarrow \exists X \rightarrow A$ かつ X が最小集合となるものである。

結合従属性 (JD) : $*[S_1, S_2, \dots, S_m]$ が関係 R で成立するとは、 R が常に次の条件を満足する場合である。

$$R = R[S_1] * R[S_2] * \dots * R[S_m]$$

上記の JD において、 $S_i \supset S_j$ となるような i, j が存在しないとき、この JD を簡約であるという。また JD で $m=2$ の場合を特に多値従属性 (MVD) と呼ぶこともある。

従属性の集合を D で表わし、 $D = F \cup J$ とする。ここで、 F は FD 集合、 J は JD 集合である。 D, F, J の要素を、

それぞれ、 α , β , γ で表わす。従属性集合 D より導出可能なすべての従属性を D の閉包と呼び D^* で表わす。 $\alpha \in D^*$ の場合、 α は D より導出可能であるといい、 $D \models \alpha$ で表わす。

3. 関係の横切りに対する基本的考察

関係の分解には大きく分けて、射影分解と横切りとがある。射影分解は、 R を $R = R_1 * R_2 * \dots * R_m$ を満足する m 個の関係に分けるものである。 R において、 $FD : X \rightarrow Y$ が成立すると、次のような射影分解ができる。

$$R = R[X, Y] * R[R - Y] \quad (i)$$

$JD : [S_1, S_2, \dots, S_m]$ が成立すると次のような射影分解が可能である。

$$R = R[S_1] * R[S_2] * \dots * R[S_m] \quad (ii)$$

1 つの FD の管理は (i) の場合 $R[X, Y]$ で管理すればよく、(ii) の場合は各関係が結合によって矢張りれる組を持たないように管理するとよい。しかし、 FD や JD が 1 個以上ある場合には、射影分解で管理するのがむづかしいことがある。このため、横切りとの併用について考察する。関係の横切りは次のように定義される。

$$R = R_1 \cup R_2 \cup \dots \cup R_m$$

ここで、すべての i と j の組合せについて $R_i \cap R_j = \emptyset$ が成立すれば、基本横切りという。次のような性質が成立する。

P1: 関係 R が FD 集合 F を満足するとき, 横切りによって得られた関係も F を満足する。もとの FFD が FFD ではない FD となったり, 新しい FD が加わることもある。

P2: 関係 R が JD 集合 J を満足するとき, 横切りによって得られた関係では J が成立するとは限らない。

基本横切りのうち次の条件を満足するものを, GROUP-BY $[X]$ と呼ぶ。

$$\forall i \quad R_i[X] = \text{定数}$$

P3: 関係 R に GROUP-BY $[X]$ を適用すると各 R_i は FD: $\phi \rightarrow X$ も満足する。FFD: $Y \rightarrow Z$ は $(Y-X) \rightarrow Z$ となる。

P4: JD: $*[S_1, S_2, \dots, S_m]$ を満足する関係 R に, GROUP-BY $[X]$ を適用すると, 各 R_i は JD': $*[S_1-X, S_2-X, \dots, S_m-X]$ を満足する。

P5: 次に示す D_1 と D_2 は, $D_1^* = D_2^*$ が成立する。

$$D_1 = \{ *[S_1, S_2, \dots, S_m], \phi \rightarrow X \}$$

$$D_2 = \{ *[S_1-X, S_2-X, \dots, S_m-X], \phi \rightarrow X \}$$

4. 横切りによる関数従属性の保持

FD: $Y \rightarrow A$ を GROUP-BY $[X]$ を適用した後に保持する方法を考える。各部分関係を R_i とし, $Z = X \cap Y$ とする。

(1) $A \in X$ の場合 $R_i[ZA] = R_j[ZA]$ となる関係
もすべてまとめて S_R とする。

$S_R[A] \neq S_R[A], S_R[Z] = S_R[Z]$ とな
る S_R と S_R は次の条件を満足する。

$$S_R[Y-X] \cap S_R[Y-X] = \emptyset$$

(2) $A \notin X$ の場合 この場合はさらに次の2つの場合に分
けられる。

(2-1) $X \subseteq Y$ の場合 この場合各部分関係で FD:
 $(Y-X) \rightarrow A$ を管理すればよい。

(2-2) $X \not\subseteq Y$ $Y \rightarrow A$ は各部分関係における管理の
みでは不十分で、 $R_i[Z]$ の等しいすべての部分関係上
で成立するかどうかの管理が必要である。

(証明) (1) まず次を証明する。「関係 R で $Y \rightarrow A$ が成立す
ることと、 R を $GROUP-BY[A]$ で横切り分解して得
えた $T_1, T_2, \dots, T_m$ において、すべての i と j ($i \neq j$) の
組合せに対して、 $T_i[Y] \cap T_j[Y] = \emptyset$ となることは等
価である。」 $T_i[Y] \cap T_j[Y] \neq \emptyset$ の場合、共通に含まれ
る Y の値を y とする。 $T_i[YA]$ は $y a_i$ を含み、 $T_j[YA]$
は $y a_j$ を含み、 $GROUP-BY$ の条件から $a_i \neq a_j$ であ
る。すなわち、 $R[YA]$ は $y a_i, y a_j$ ($a_i \neq a_j$) という
組を含むため $Y \rightarrow A$ に反する。逆に条件が満足されればこの

ようなことはない。

上記の $S_1, S_2, \dots$ は $GROUP-BY [Z A]$ を行ってできた関係となる。 $S_R[A] = a_R, S_R[A] = a_R, S_R[Z] = S_R[Z] = \text{ある}$ とすると, $S_R[Z A] = \text{ある } a_R, S_R[Z A] = \text{ある } a_R$ となる。上記と同様にして, この場合は R で $Y \rightarrow A$ が成立するのと, 次の条件は等価となる。

$$S_R[Y - X] \cap S_R[Y - X] = \phi.$$

(2-1) $X \leq Y$ かつ $A \notin X$ の場合, 各部分関係において $R_i[X] \cap R_j[X] = \phi$ が成立するため, $(Y - X) \rightarrow A$ が各部分関係で成立することと, R で $Y \rightarrow A$ が成立することは等価になる。

(2-2) $X \not\leq Y$ $A \notin X$ の場合, まず $Z = X \cap Y \neq \phi$ の場合も考える。 $R_i[Z]$ の値の等しい部分関係をすべてまとめると等価的に $GROUP-BY [Z]$ をしたことになる, $Z \leq Y$ であるため (2-1) と同じ方法で管理できる。 $Z = \phi$ の場合は $X \cap Y A = \phi$ となり, $GROUP-BY$ に使われた X はこの FD と全く共通属性を持たない。したがって, 元の R にもとじて FD の管理を行わなければならない。 $Z = \phi$ の場合, すべての $R_i[Z]$ は等しいと定義すれば上記の特殊な場合を含めることができる。 (証明終り)

与えられた FD に対して, 管理が容易になるように X を選

び、 $GROUP - BY [X]$ をほどこすことになる。管理の容易なのは次の場合である。

a. $Y \rightarrow A$ に対し、 $X \subseteq Y$ である X で $GROUP - BY [X]$ を行なう (上記の (2-1))。

b. $Y \rightarrow A$ に対し、 $GROUP - BY [A]$ を適用する。
(上記 (1) の特殊な場合)。

P6: すべての $FD: Y_i \rightarrow A_i$ を部分関係における FD 管理 (上記 a) のみで保持できるための必要十分条件は、
 $\cap Y_i \neq \emptyset$ であることである。この場合、 $X = \cap Y_i$ によって
 $GROUP - BY [X]$ を適用するとよい。

[例] $R (R = ABCD)$ で、2つの $FD: AB \rightarrow D$ と
 $BC \rightarrow D$ が成立しているとする。射影分解による FD の管理
では、 $R[ABD]$ 、 $R[BCD]$ 、 $R[ABC]$ の3つの関
係を作り、前の2つで2つの FD を管理することになる。3
番目の関係はすべての関係を結合して R もどすために必要
となる。横切り分解では次のようになる。

まず、 $GROUP - BY [B]$ をほどこし、 $R_1, R_2, \dots, R_m$ を
得る。P6により、各 R_i で $A \rightarrow D$ と $C \rightarrow D$ を管理するとよ
い。これは、各 R_i にさらに $GROUP - BY [D]$ をほどこ
すと上記 b の場合となり、各 R_i を分けた部分関係で A と
 C の値集合が互いに素になるようにすればよい。

5. 関数従属性によるキー破壊の扱い

FDによるキー破壊のある場合は、射影分解のみではFDの保持の出来ないことが知られている。本節では、著者ら⁽²⁾とFurtado⁽⁴⁾によって独立に与えられた場合を含めた一般的なキー破壊の扱いについて考察する。

関係 R ($R = ABC$) において、次の2つのFD f_1 と f_2 が成立している場合を考える。

$$f_1: AB \rightarrow C \qquad f_2: C \rightarrow B$$

(方法1) GROUP-BY [B] を適用する。

各 R_i では次の2つのFDが成立する。 $A \rightarrow C \quad \phi \rightarrow B$

f_1 は a の場合に対応するので各部分関係で $A \rightarrow C$ を管理するだけで保持できる。 f_2 は b の場合に対応しているので、各 R_i において $R_i[C]$ の値が互いに素という条件で管理できる。

(方法2) GROUP-BY [A] を適用する。

各 R_i では次の2つのFDが成立する。 $\phi \rightarrow A \quad B \leftrightarrow C$

f_1 は a の場合に対応するので各部分関係で $B \rightarrow C$ を管理すればよい。 f_2 は a でも b でもないため、 R に対する管理を必要とする。

(方法3) GROUP-BY [C] を適用する。

各 R_i では $\phi \rightarrow BC$ のみが成立する。 f_1 は b の場合に相当し

ているため、各部分関係で AB の値が互いに素になるよう管理すればよい。 f_2 は a に対応しているため、各部分関係で $\phi \rightarrow B$ が成立するように管理する。すなわち、各部分関係において B の値は定数である。

(方法4) 2つの属性による $GROUP-BY$ の適用
 $GROUP-BY[BC]$ は結果として方法3と同じになる。
 $GROUP-BY[AB]$ は f_1 により $GROUP-BY[ABC]$ と同等となり、関係を一組づつの部分関係に分ける操作となり無意味である。

$GROUP-BY[AC]$ は、 f_1 に対しては (1) で $\emptyset \neq \phi$ の場合となり、 f_2 に対しては (2-2) となるため、両方とも簡単には管理できない。

6. 横切りによる結合従属性の保持

本節では、 $JD: * [S_1, \dots, S_m]$ と $GROUP-BY[X]$ との関係について考察する。 $* [S_1, \dots, S_m] \models * [S_i, Y_i]$ ($Y_i = \bigcup_{j \neq i} S_j$) であるため、まず $* [S_i, Y_i]$ の保持について考察する。

P7: R に対して $GROUP-BY[X]$ を適用してできた関係を $R_1, R_2, \dots, R_n$ とする。 R で $* [S_i, Y]$ が成立することと、各 R_i で $* [S_i, Y]$ が成立することが等価となるための必要十分条件は、 $S_i \cap Y \supseteq X$ が成立することである。

(証明) X が $S_i \cap Y$ にない属性も含めば、同じ $S_i \cap Y$ の値が複数の R_i に出現することになる。したがって、各 R_i で $*[S_i, Y]$ が成立しないことがある。 $X = S_i \cap Y$ であれば、各 R_i で、 $*[S_i - X, Y - X]$ と $\phi \rightarrow X$ が成立する。前者は $(S_i - X) \cap (Y - X) = \phi$ であるため直積である。P4により各 R_i で $*[S_i, Y]$ が成立していることになる。 $S_i \cap Y \neq X$ の場合は、各関係は上記の部分関係の和集合となるが、やはり $*[S_i, Y]$ も満足する。GROUP-BY をしない場合に相当する $X = \phi$ の場合も含まれている。(証明終り)

P8: R に対して GROUP-BY $[X]$ も適用してできた関係も $R_1, R_2, \dots, R_n$ とする。 R で $j: *[S_1, S_2, \dots, S_m]$ が成立することと、各 R_i で同じ j が成立することとが等価になるための必要十分条件は、 $\cap S_i \supseteq X$ となることである。

(証明) P7 が $Y = Y_1, Y_2, \dots, Y_m$ のすべての場合適用できる条件を求めると $\cap S_i \supseteq X$ となる。 $\cap S_i \supseteq X$ の場合の等価性の証明は P7 の場合とほぼ同じである。(証明終り)

P9: P8 で、FD: $Y \rightarrow X$ が成立する場合は、条件は $\cap S_i \supseteq Y$ となる。

(証明) GROUP-BY $[X]$ で得られた部分関係では、 Y の値が互いに素である。したがって、各 R_i で j が成立するのと R で j が成立するのは等価になる。(証明終り)

P8 は横切り分解によって JD が完全に保持できる条件を示しているが非常に制限が強い。P9 を用いると巡回 JD も扱うことができる。

[例1] JD j : $*[AB, AC, ADE, EF, EGH, HI]$ を考える。JD の各成分を枝に、各属性を節点に対応させたハイパーグラフ表現は図1に示されている。

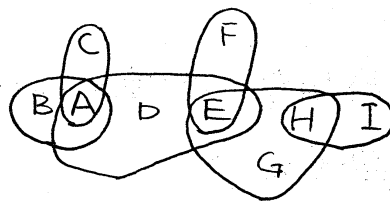


図1

ここで、 $A \rightarrow H$, $E \rightarrow H$ という FD が成立すれば、

GROUP-BY [H] によって、次の JD は部分関係における管理で保持できる。

$*[AB, AC, ADEFGHI]$

$A \rightarrow H$ による。

$*[EABCD, EF, EGH I]$

$E \rightarrow H$ による。

$*[HABCDEFG, HI]$

GROUP-BY [H] による。

j はこれらの JD より導かれる。この場合、FD の保持のため、各部分関係における A や E の値集合が互いに素になるように管理しなければならない。

[例2] JD j : $*[ABC, CDE, AEF]$ に対応するハイパーグラフは図2に示されている。ここで

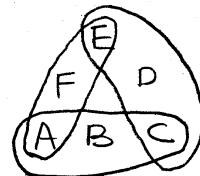


図2

$A \rightarrow E$, $C \rightarrow E$ という FD が成立していれば、

GROUP-BY [E] によってできた部分関係で同じ JD を管理すると JD の保持ができる。

7. むすび

本稿では、従属性の保持と横切り分解に関して基本的な考察を行った。単なる GROUP-BY では能力が限られているため、より一般的な横切り分解を考えるのが今後の課題である。

謝辞 御指導，御検討いただいた京都大学工学部矢島脩三教授に深謝します。

文献

- (1) E.Sciore, "Improving Database Schemes by Adding Attributes," ACM Conference on Principles of Database Systems, pp.379-383 (1983).
- (2) Y.Kambayashi, K.Tanaka, K.Takeda and S.Yajima, "Representation of Relations for Database Output Utilizing Data Dependencies, Hawaii International Conf. System Sciences, pp.69-78, (1982).
- (3) Y.Kambayashi and M.Yoshikawa, "Query Processing Utilizing Dependencies and Horizontal Decomposition," ACM SIGMOD International Conference on Management of Data, (1983).
- (4) A.L.Furtado, "Horizontal Decomposition to Improve a non-BCNF Scheme," ACM SIGMOD Record 12,1, pp.26-32 (1981).