

Extended Conflict-free Set 再考

勝野 裕文 (Hirofumi Katsuno)
(電電公社 武蔵野通研)

1. はじめに

データベースの分野ではデータ従属性に関する理論が盛んに研究されている。ここでいうデータ従属性の定義は広い意味から狭い意味までさまざまに使われている。しかし、関数従属 (FD), 多値従属 (MVD), 結合従属 (JD) がデータ従属性の基本概念であることに対する異論はない。ここで、FD, MVD, JD は次のような直観的な意味を持つ。

FD — 関数関係 (表に依存しない概念)

MVD — 表 r を射影によって r_1 と r_2 に分解した時に、
 $r = r_1 \bowtie r_2$ となる為の必要十分条件 (\bowtie は結合を示す)

JD — 表が射影によって n 個に分解 ($r = r_1 \bowtie r_2 \bowtie \dots \bowtie r_n$)
できる為の必要十分条件

このように一つの MVD 又は JD は表の世界において明確な意味を持つ。しかし、表の世界と離れた現実世界における

MVD, JDの意味, 特にMVD, JDからなる集合の意味が不明確な場合がある。意味の不明確なMVD集合を排除する為に, conflict-freeと呼ばれるMVD集合のクラスが考えられている。^[8]

我々は conflict-free MVD集合にFDの概念をとりこむ為に, extended conflict-free という概念を既に提案した。^[6] 本稿では, その extended conflict-free の直観的な意味とその問題点について述べる。尚, 関係データベースの基本概念, 用語は既知であると仮定する。本稿で未定義な概念, 用語については [9] を参照すること。

2. MVD, JDの現実世界における意味

一つのMVD, JDに注目すると, その現実世界における意味はかなり明確になる。初めに, MVD, JDを定義する。

[定義]

属性集合 U の部分集合 X_1, \dots, X_n が $U = \bigcup_{i=1}^n X_i$ を満たす時に, 関係 $r(U)$ において, 結合従属 (JD) $\bowtie [X_1, \dots, X_n]$ が成り立つとは,

$$r(U) = r[X_1] \bowtie \dots \bowtie r[X_n]$$

が成り立つことをいう。但し, $r[X_i]$ は $r(U)$ の X_i 上への射影を示す。

上の定義において、 $n=2$ の場合、 $r(U)$ においてMVDが成り立つと(11)、 $\bowtie[X_1, X_2]$ のかわりに、

$X_1 \cap X_2 \twoheadrightarrow X_1 - (X_1 \cap X_2)$ (又は $X_1 \cap X_2 \twoheadrightarrow X_2 - (X_1 \cap X_2)$) と書く。尚、左辺が共通なMVDはまとめて、 $X \twoheadrightarrow Y_1, \dots, X \twoheadrightarrow Y_n$ を $X \twoheadrightarrow Y_1 | \dots | Y_n$ と書く。

一つのJD(MVD)は次の場合に成り立つ。

(*) 属性集合 X_i に関する一つの関連が存在して、その関連のデータが関係 $r_i(X_i)$ で表わされている時、

$$r = r_1(X_1) \bowtie \dots \bowtie r_n(X_n)$$

とすると、 r においてJD: $\bowtie[X_1, \dots, X_n]$ が成り立つ。

[例1]

学生(S)がどの講義(C)を履習しているかを示す関連に関するデータを $r_1(S, C)$ で表わし、先生(T)がどの講義(C)を教えているかを $r_2(T, C)$ で、講義(C)でどの本(B)が教材となっているかを $r_3(B, C)$ で、学生(S)の専攻(M)と学年(Y)を $r_4(S, M, Y)$ でそれぞれ表わすとする。この時、 $r = r_1 \bowtie r_2 \bowtie r_3 \bowtie r_4$ とすると、 r において $\bowtie[SC, TC, BC, SMY]$ が成り立つ。

従って、 r においてJD: $\bowtie[X_1, \dots, X_n]$ が成り立つことの

つの意味は、異なる関連が m 個あって、 r はそれらの関連のデータを一つの関係で表わしているという点である。しかし (★) で構成した r においては、

(GC) 任意の i に対して、 $r[X_i] = r_i$ (GC: globally consistent) が成り立つとは限らない。即ち、(★) のように r を構成すると情報が失われる可能性がある。一般に、pairwise consistent を

(PC) 任意の i, j ($i \neq j$) に対して $r_i[X_i \cap X_j] = r_j[X_i \cap X_j]$ とすると、 $n=2$ であれば、(PC) が成り立てば、常に (GC) が成立する。一方、 $n \geq 3$ の時は、(PC) が成り立ったとしても、(GC) が成り立つとは限らない。

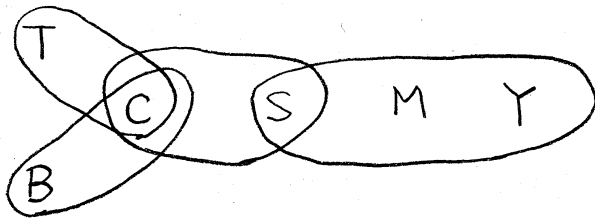
JD と MVD との間の関係は単に MVD が JD の特殊な場合というだけでなく、一つの JD が MVD の集合と等価な場合がある。これが、MVD の集合の一つの意味である。

ある JD が導ける MVD の集合は次のように規定される。但し、“導ける”とは、 $\forall R$ が成り立つすべての関係において、 $X \Rightarrow Y$ が成り立つ時、 $\forall R$ から $X \Rightarrow Y$ が導けるという。

$R = \{X_1, \dots, X_n\}$ に対して、 $\forall R$ によって $\forall [X_1, \dots, X_n]$ という JD を示すことにする。 $\forall R$ に対応するハイパー・グラフとは各 X_i の要素を頂点、 X_i を辺とするハイパー・グラフをいう。

[例2]

例1の JD に対応するハイパー・グラフを下に示す。



定理 1 [5]

$\mathcal{M}R$ から $X \Rightarrow Y$ が導ける。

\Leftrightarrow $\mathcal{M}R$ に対応するハイパー・グラフにおいて、 X を除去して得られるグラフの連結成分を Y_1, \dots, Y_m とすると、ある i_1, \dots, i_k が存在して、 $Y = Y_{i_1} \vee \dots \vee Y_{i_k}$ とかける。

[例] 3]

例 1 の JD; $\mathcal{M}[SC, TC, BC, SMY]$ から例 2 は、

$CS \Rightarrow T | B | MY, \quad S \Rightarrow CTB | MY$

が導ける。

逆に、MVD の集合から JD が導ける場合、即ち MVD の集合と JD が等価になる場合は次の節で示す。

3. Conflict-free MVD 集合

一つの MVD, JD の意味は 2 節でみたように比較的わかり易い。しかし MVD からなるある種の集合 (MVD の推移律が適用できたり、MVD の左辺が他の MVD によって分割されるような集合) には意味的な問題点があることが示されている。^{[7][8]} このような意味的な問題点が生じない MVD の集合の

クラスとして, conflict-free という概念が定義されている。^[8]

[定義]

Γ を FD と MVD からなる集合とする時, $DEP(X)$ が次の条件を満たす時, $DEP(X)$ を Γ における X の従属基であるという。

$$(1) X = U \text{ ならば } DEP(X) = \{\phi\}$$

$$(2) X \text{ の } \Gamma \text{ による閉包を } X^* \text{ とする時, } X^* = U \text{ ならば,}$$

$$DEP(X) = \{A_1, \dots, A_m \parallel \phi\} \quad \text{かつ}$$

$$U - X = \{A_1, \dots, A_m\}$$

(1), (2) 以外の場合は次の (3) ~ (6) が成り立つ。

$$(3) DEP(X) = \{A_1, \dots, A_m \parallel Y_1, \dots, Y_k\} \text{ であり, } \{A_1, \dots, A_m, Y_1, \dots, Y_k\} \text{ は } U - X \text{ の分割である。}$$

$$(4) \Gamma \text{ から } X \rightarrow A_i \text{ (} 1 \leq i \leq m \text{)} \text{ が導けるが, } B \in Y_1 \cup \dots \cup Y_k \text{ であれば, } \Gamma \text{ から } X \rightarrow B \text{ は導けない。即ち, } X^* = X \cup \{A_1, \dots, A_m\}$$

$$(5) \Gamma \text{ から } X \Rightarrow Y_i \text{ (} 1 \leq i \leq k \text{)} \text{ が導ける。}$$

$$(6) Y \text{ の任意の空でない部分集合 } Y' \text{ に対して, } \Gamma \text{ から } X \Rightarrow Y' \text{ は導けない。}$$

尚, Γ が MVD のみからなる場合又は $X^* = X$ の場合は, $DEP(X)$ を単に $\{Y_1, \dots, Y_k\}$ と書く。

[定義]

MVD からなる集合を M とし, X, Y をそれぞれ M のある M

VD の左辺であるとする時、 X と Y が conflict-free であるとは、次の二条件が成り立つことをいう。

$$\begin{aligned} \textcircled{1} \quad \text{DEP}(X) &= \{V_1, \dots, V_k, X_1, \dots, X_i, Z_a^U Y_1, \dots, Y_j\} \\ \text{DEP}(Y) &= \{V_1, \dots, V_k, Y_1, \dots, Y_j, Z_b^U X_1, \dots, X_i\} \\ &\text{かつ } Z_a^U X = Z_b^U Y \text{ とかける。} \end{aligned}$$

$$\textcircled{2} \quad \{V_1, \dots, V_k\} \subset \text{DEP}(X \cap Y)$$

M 内のMVDの左辺の任意の pair, X, Y が conflict-free である時、 M が conflict-free であるという。

JD: $\mathcal{M}R$ に対して、ハイパーグラフを考え、そのグラフにおいて、サイクルの有無を議論することができ。ハイパーグラフがサイクルを持たない時に、 $\mathcal{M}R$ を acyclic JD と呼び、conflict-free MVD 集合との間に次の定理が成り立つ。
定理2^[2]

M が conflict-free MVD 集合である。

$\Leftrightarrow M$ と同等な acyclic JD が存在する。

ここでいうハイパーグラフのサイクルの概念の定義は面倒なので、省略するが、ここでは acyclic JD を特徴づける次の定理を示す。

定理3^[2]

$\times R$ が acyclic JD である。

\Leftrightarrow ハイパー・グラフ R から導ける MVD の集合と $\times R$ が同値である。

\Leftrightarrow (PC) が成り立てば、(GC) が成り立つ。

定理 3 で示した他にも、acyclic JD (又は acyclic hypergraph) がいくつかの望ましい特徴を持つことが示されている。

4. Extended Conflict-free

前節で定義した conflict-free は MVD の集合に対する概念であって、FD の存在は特に議論されていない。[6] では、FD と MVD を一緒に考えて、extended conflict-free という概念を次のように定義した。

[定義]

Γ を FD と MVD からなる集合とする時、 X の従属基が

$$DEP(X) = \{A_1, \dots, A_m \parallel Y_1, \dots, Y_k\}$$

であれば、 X の退化 MVD 集合 M_X を

$$M_X = \{X^* \twoheadrightarrow Y_1, \dots, X^* \twoheadrightarrow Y_k\} \quad \text{但し } X^* = X \setminus \{A_1, \dots, A_m\}$$

と定義する。 Γ の退化 MVD 集合 M を

$$M = \bigcup_X M_X \quad (\text{但し } \bigcup_X \text{ において、} X \text{ は } \Gamma \text{ の FD 又は MVD の左辺})$$

と定義する。

[定義]

Γ の退化 MVD 集合 M が conflict-free である時、 Γ が extended conflict-free であるという。

conflict-free MVD 集合が acyclic JD と同値であるように、extended conflict-free 集合も JD と次のような対応がある。

[定義]

$X \rightarrow Y$ が $\bowtie[X_1, \dots, X_n]$ と適合しているとは、ある i に対して、 $X \cup Y \subset X_i$ が成り立つことをいう。FD からなる集合 F の任意の元が $\bowtie[X_1, \dots, X_n]$ と適合する時に、 F は $\bowtie[X_1, \dots, X_n]$ に適合しているという。

[定義]

F が $\bowtie[X_1, \dots, X_n]$ に完全に適合しているとは、 F が $\bowtie[X_1, \dots, X_n]$ に適合していて、かつ $X \rightarrow Y$ が F の元であり、 $X \subset X_i$ ならば、 $Y \subset X_i$ が成り立つことをいう。

[例4]

例1において、 $F = \{C \rightarrow T, S \rightarrow MY\}$ とすると、 F は $\bowtie[SC, TC, BC, SMY]$ に適合しているが、完全には適合していない。

定理4

FD と MVD からなる集合 Γ が extended conflict-free である。

\Leftrightarrow ある FD の集合 F と acyclic JD: $\bowtie R$ が存在して、 F は $\bowtie R$ と適合していて、かつ $F \cup \{\bowtie R\}$ は ρ と同値である。

\Leftrightarrow ある FD の集合 F と acyclic JD: $\bowtie R$ が存在して、 F は $\bowtie R$ と完全に適合していて、かつ $F \cup \{\bowtie R\}$ は ρ と同値である。

extended conflict-free という概念を定義する目的の一つは、現実世界の自然な制約をデータ従属性の集合で記述する時に、その集合がどのように特徴づけられるかを明らかなにすることであった。その目的がらみると、extended conflict-free には次の例に示すような問題点がある。即ち、現実世界の自然な制約で、extended conflict-free 集合に対応しない場合がある。

[例 5] ^[5]

学生 (S) がどの講義 (C) を履習しているかを示すデータに対応する関係 $r_1(S, C)$ と、講義 (C) が何時 (H) にどの教室 (R) で行なわれるかを示す関係 $r_2(C, H, R)$ がある時に、データ従属性の集合 ρ として、

$$\rho = \{SH \rightarrow CR, RH \rightarrow C, CH \rightarrow R, \bowtie [CS, CRH]\}$$

を考える。この ρ に対応する現実世界の制約は自然なものであるが、 ρ と同値な extended conflict-free 集合は存在しない。

extended conflict-freeはこのような問題点を含むが、望ましい性質も持っている。それは、制約を extended conflict-freeに限ると、次に示すような忠実な分解が存在するという点である。

[定義]

ρ をデータ従属性の集合とする時、 ρ を制約として持つ、属性集合 R 上の関係スキーマを $\langle R, \rho \rangle$ と書くことにする。

[定義]

次の (1)~(4) の性質を持つ、関係スキーマの集合 $\{\langle R_1, F_1 \rangle, \dots, \langle R_n, F_n \rangle\}$ を $\langle R, \rho \rangle$ の忠実な分解という。

(1) 各 F_i ($1 \leq i \leq n$) は FD の集合である。

(2) ρ と $F_1 \cup \dots \cup F_n \cup \{\forall [R_1, \dots, R_n]\}$ は同値である。

(3) ρ が導けるすべての FD の集合を F とし、 F の R_i 上への制限を $F|_{R_i}$ とすると、 $F|_{R_i} = (F_i)^+$ が成り立つ。但し、 $(\cdot)^+$ は FD 集合の閉包を示す。

(4) 各 $\langle R_i, F_i \rangle$ は第3正規形である。

[注意]

上の条件(2)は次の(2)'と同値であることが知られている。^[3]

(2)' $\text{SAT}(F_1, \dots, F_n) = \{\langle r_1, \dots, r_n \rangle \mid r_i \text{ は } F_i \text{ が成り立つ関係であり, } r_1, \dots, r_n \text{ で (PC) が成り立つ}\}$ とし、

$\text{SAT}(\rho) = \{r \mid r \text{ は } \rho \text{ が成り立つ関係}\}$,

$r \in \text{SAT}(\Gamma)$ に対して $\Delta(r) = \langle r[R_1], \dots, r[R_n] \rangle$ とすると, Δ は $\text{SAT}(\Gamma)$ から $\text{SAT}(F_1, \dots, F_n)$ への全単射であり, Δ^{-1} は結合である.

定理 5

Γ が extended conflict-free ならば, $\langle R, \Gamma \rangle$ の忠実な分解が存在する.

extended conflict-free が忠実な分解に対応しているか, 即ち $\langle R, \Gamma \rangle$ の忠実な分解が存在すれば, Γ が extended conflict-free 集合に同値になるかという点, それだけ成立しない. 忠実な分解は, FD集合とJDの適合性に関係している.

定理 6

Γ を $\{\text{MR}\} \cup F$ とする時, F が MR と適合しているならば, $\langle R, \Gamma \rangle$ の忠実な分解が存在する. 逆に, $\langle R, \Gamma \rangle$ の忠実な分解が存在すれば, ある JD; MR と FD集合 F で, $\{\text{MR}\} \cup F$ が Γ と同値で, F が MR と適合しているものが存在する.

[注意]

定理 4 と定理 6 から定理 5 は容易に導かれる.

以上のことから考えると直観的に

extended conflict-free \cong acyclic + 忠実な分解

という見方をすることができるといえる。

5. おわりに

extended conflict-free では、[1] での acyclic という概念を前提としているが、ハイパーグラフのサイクルの概念は多様なので、[1] のサイクルと異なる意味のサイクル^[4]を対象とした時の考察も必要だと考える。

参考文献

- [1] C. Beeri, R. Fagin, D. Maier, A. Mendelzon, J. Ullman and M. Yannakakis: Properties of Acyclic Database Schemes, Proc. of ACM-STOC, pp.355-362, 1981
- [2] C. Beeri, R. Fagin, D. Maier and M. Yannakakis: On the Desirability of Acyclic Database Schemes, IBM Research Report, RJ3131, 1981
- [3] C. Beeri and J. Rissanen: Faithful Representations of Relational Database Schemes, IBM Research Report, RJ2722, 1980
- [4] Fagin, R.: Types of Acyclicity for Hypergraphs and Reltional Database Schemes, IBM Research Report, RJ3330, 1981
- [5] Fagin, R., A.O. Mendelzon, and J.D. Ullman: A Simplified Universal Relation Assumption and its Properties, ACMTODS, Vol.7, No.3, pp.343-360, 1982
- [6] Katsuno, H.: A Desirable Class of Sets of Functional Dependencies and Multivalued Dependencies, AL81-65, 1981
- [7] Kambayashi, Y., K. Tanaka, and S. Yajima: Semantic Aspects of Data Dependencies and their Application to Relational Database Design, Proc. of COMPSAC 79, pp.398-403, 1979
- [8] E. Sciore: Real-World MVDs, Proc. of ACM-SIGMOD, pp.121-132, 1981
- [9] J.D. Ullman: Principles of Database Systems, Computer Science Press, 1980