数理解析研究所講究録
第495巻 1983年 27-50



# ネットワークモデルによる制約表現

京大工学部　上林弥彦　(Yahiko Kambayashi)
京大工学部　古川哲也　(Tetsuya Furukawa)
京大工学部　矢島脩三　(Shuzo Yajima)

## 1. まえがき

データベースシステムの主要な目的は、個々の利用者の持つデータをまとめることにより、冗長度を減らし全体のデータを統一的に管理できる等、データ共有による利点を生かす点にある。E.F.Codd によって提唱された関係データベースは、使い易く将来システムであると考えられ、その理論は数学的にも扱い易いために大きく発展してきた。このモデルはスキーマの変更などの自由度も大きい反面、効率が悪いという欠点もある。これに対し、ネットワークデータベースは質問が作りにくく、使い易さの点で関係データベースに劣るが効率が非常によく、予め質問が定まっている日常業務的なシステムでは問題のないため、大規模データベースでは今後も使用されると考えられる。また、効率の良さという長所を生かし使いにくいという短所を避けるため、利用者が関係データベースのように使える関係インタフェースや使い易い補助システムの開発の研究も多く、これらは今後の重要な方向であ

る。これらの研究のためには、ネットワークモデルに対するスキーマ設計、ネットワークモデルの扱える意味制約、関係モデルとネットワークモデルの対応 といった問題を数学的に明確にしなければならない。

関係モデルはデータの構造と意味制約を独立に扱えるが、ネットワークモデルはデータの構造そのものに制約があり、構造と意味制約を切り離して考えることはできない。そのため、2つのモデル間の対応やネットワークデータベースの設計を議論する際には、ネットワークモデルの構造による制約を明確にする必要があるが、それに関する研究はほとんど行なわれていない。Lienは関係モデルとネットワークモデルの対応について空値の扱いも含めて研究を行なっているが、これは関係モデルの立場からの見方による所が多く、2つのモデルが扱える意味制約の違いまで言及していない。ネットワークデータベースのスキーマ設計についても体系的な方法はなく、実世界を実体関連モデルで表現し、その対応関係からネットワークモデルを設計するという手法が多く用いられ経験的な直感による所が多い。

本稿では、CODASYL型のネットワークデータベースを対象とし、関係データベース理論をネットワークモデルに適用して2つのモデルで扱える意味制約の違いを空値に関する問

題を含めて明確にすると共に、その結果を踏まえたネットワークデータベースの設計手法を与える。また、ネットワークモデルの立場から見たLienの研究の問題点も指摘している。

分散データベースにおける異なるデータベースシステム間の通信やデータベース管理システムの変更によるスキーマの変換など、異なるモデルのデータベースの対応を明確にしなければならない場合がある。本稿の結果は、これらの問題に対しても広く応用することができる。

## 2. 基本的事項

n個の集合$D_1, D_2, \cdots, D_n$（必ずしも異なるとは限らない）の要素$d_1, d_2, \cdots, d_n$の順序づけられた組$(d_1, d_2, \cdots, d_n)$の全体からなる集合を直積(cartesian product)といい、直積の部分集合Rを関係(relation)という。$D_1, D_2, \cdots, D_n$をRの定義域(domain)といい関係Rの各要素を組(tuple)という。また、関係Rを構成する定義域の識別のために、一意な名前$A_1, A_2, \cdots, A_n$を与え属性と呼ぶ。

関係に対して行なわれる操作に関係代数(relational algebra)があり、次にその主要な操作を示す。組rの属性の集合Xの値をr[X]で表わすと

射影　関係Rの属性集合Xについての射影をR[X]とし、R[X]

$=\{r[X]: r\in R\}$ と定義する。

**選択**　$\theta$ を比較演算子 $\neq, =, >, \geq, \leq, <$ の1つとするとき、関係 $R$ の属性集合 $X$ についての制約を定数'c'に対し、$R[X\theta 'c'] = \{r: r\in R, r[X]\theta 'c'\}$ と定義する。ただし $r[X]$ と'c'は $\theta$ に関して比較可能(domain compatible)であるものとする。

**(自然)結合**　関係 $R_1(X,Y)$ と $R_2(X,Z)$ で $Y\cap Z=\phi$ のときの自然結合を $R(X,Y,Z)=R_1 * R_2$ とし、次のように定義する。$R(X,Y,Z)=\{r: r[X,Y]\in R_1, r[X,Z]\in R_2\}$

　$m$ 個の関係 $R_1(X_1), R_2(X_2), \cdots, R_m(X_m)$ に対し、$R_i=R[X_i]$ を満足する関係 $R$ が常に存在するという仮定を**普遍関係仮定**(universal relation assumption)といい、$R$ を普遍関係という。

　関係 $R(U)$ の組 $r, s$ と属性集合 $X, Y$ について、任意の $r, s$ に対し $r[X]=s[X]$ ならば $r[Y]=s[Y]$ が成り立つとき、$X$ から $Y$ に**関数従属性**(functional dependency: FD)があるといい、$X\rightarrow Y$ で表わす。$X$ をFDのキーという。また、$Z=U-X-Y$ としたとき、任意の組 $r$ について $(R[XZ=r[XZ]])[Y]=(R[X=r[X]])[Y]$ を満たすときに、$X$ から $Y$ に**多値従属性**(multivalued dependency: MVD)があるといい、$X\twoheadrightarrow Y$ で表わす。$X\twoheadrightarrow Y_1, X\twoheadrightarrow Y_2, \cdots, X\twoheadrightarrow Y_n$ が成り立つ場合は、$X\twoheadrightarrow Y_1|Y_2|\cdots|Y_n$ と表わし、$X$ をMVDのキー、各 $Y_i$ が $U-X$ の極小

分割のとき<u>従属基</u>といいその集合をDEP(X)で表わす。MVD集合が<u>無矛盾</u>(conflict-free)であるとは、任意の2つのMVDキーX,Yに対し、

$DEP(X)=\{V_1,V_2,\cdots,V_k,X_1,X_2,\cdots,X_i,(Z_aY_1Y_2\cdots Y_j)\}$

$DEP(Y)=\{V_1,V_2,\cdots,V_k,Y_1,Y_2,\cdots,Y_j,(Z_bX_1X_2\cdots X_i)\}$

$i\geq 0, j\geq 0, k\geq 0$　$\{V_1,V_2,\cdots,V_k\}\subseteq DEP(X\cap Y)$ かつ $Z_aX=Z_bY$

を満たすときである。MVD集合が<u>無衝突</u>(contention-free)であるとは、任意の2つのMVDキーX,Yに対し、

$X\cap Y=X$ or $Y$ or $\emptyset$　又は

$X\cap Y$ は $ZW_1\supset X, ZW_2\supset Y$ であるような異なる従属基 $W_1,W_2$ を持つキーZを含む。

を満たすときである。関係R(X)、$X=X_1\cup X_2\cup\cdots\cup X_m$ に対し、$R=R[X_1]*R[X_2]*\cdots*R[X_m]$ を満足するとき、Rは<u>結合従属性</u>(join dependency:JD) $*[X_1,X_2,\cdots,X_m]$ を満足するという。JDはハイパーグラフ(N,E)で表現することができる。ここで、Nは各属性に対応する節点の集合、Eは $X_1,X_2,\cdots,X_m$ に対応する枝の集合である。JDに対応するハイパーグラフが非巡回のとき、そのJDを<u>非巡回JD</u>(acyclic JD)という。

tを関係R(X)の組、属性集合YをXの部分集合とする。Yのすべての属性Aに対しtのAの値t[A]が空値でないとき、tは<u>Y全値</u>であるといい、すべての組がY全値のとき関係R

はY全値であるという。Y=X のとき関係Rは全値であるという。Yのすべての属性Aに対し$t$[A]が空値となるものがあるときY部分値、すべてが空値のときY空値であるという。関係RはY全値でないときY部分値、全値でないとき部分値であるという。$t$の値が空でないような属性全体の集合を$t$の対象(object)といい、obj($t$)で表わす。$Y_1, Y_2, \cdots, Y_n$をXの部分集合としたとき、関係Rが対象集合$\{Y_1, Y_2, \cdots, Y_n\}$を満足するのは、Rのすべての組$t$に対し、$\exists Y_i$, obj($t$)$=Y_i$ が満足されるときである。

ネットワークモデルにおけるレコード型とは、同じ属性集合からなるレコードの集合である。あるレコード型(親レコード型)の1つのレコード(親レコード)に別のレコード型(子レコード型)の任意個数のレコード(子レコード)を関連づけて全体に名前をつけたものを親子集合(set)といい、親レコード型のすべてのレコードに対してつくられる。これを個々のレコードではなくレコード型の関係として記述したものを親子集合型といい、S$\langle R_1, R_2\rangle$ で表わす。Sは親子集合型の名、$R_1, R_2$はそれぞれ親レコード型、子レコード型である。ネットワークモデルの構造は、バックマン線図と呼ばれる有向グラフ($N$, E)で表わされる。$N$は各レコード型に対応する節点集合、Eは各親子集合型に対応し、親レコード型に対応する節点から

子レコード型に対応する節点へ向かう有向枝の集合であり、親子集合型名のラベルをつけることもある。

## 3．ネットワーク構造がもつ制約

### 3．1　親子集合の表現する従属性

ネットワークモデルの構造は親子集合が基本単位となっている。従って、ネットワークモデルの構造に、どのような制約があるのかを議論するためには、まずその基礎となる親子集合がどのような制約を表現しているのかを明確にしておかなければならない。本節では、親子集合が表現する制約を関係データベース理論と対比させ、それがどのようにネットワークモデルで使用されているかを示す。

ネットワークモデルでは、レコードに対して次の3つの制約がある。

<u>制約1</u>：すべてのレコードは、そのレコードが子レコードとなる少なくとも1つの親子集合に参加しなければならない。

<u>制約2</u>：1つの親子集合型では、子レコードは最大1個までの親レコードを持ってよく、親レコードは任意個の子レコードを持つ。

<u>制約3</u>：1つの親子集合型で、親レコード型と子レコード型が同じ属性を持つならば、親レコードと対応する子レコード

のその属性の値は一致しなければならない。

制約1は存在制約と言われるものであるが、システムを親とする親子集合型の指定も可能であることから必然的なものではない。制約2は親レコード型と子レコード型の対応関係が1対多である、つまり子レコード型から親レコード型へのFDがあることを示している。親子集合に参加しない子レコード、子レコードを持たない親レコードも存在しえるが、このような場合については5節で検討する。制約3は構造制約と呼ばれるもので、データアクセスの効率向上のため用いられるが本来子レコードの持つ属性値は冗長なものであると言える。親レコードと子レコードで属性の値が一致しないときは属性名の変更や新しくその属性からなるレコード型をつくることで解決できる。図1(a)で属性Cの値が親レコードと子レコードで一致しなければ、属性名を変更し(b)としたり、属性Cを1つのレコード型として(c)とする。

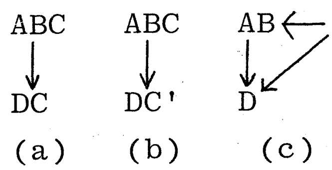



図 1

このようにレコードに関しては制約2のみを考えればよく、それは親子集合型をFDを表わしていることを示している。2つのレコード型のレコードの対応が1対多であれば、これは1つの親子集合型として表現できるが、多対多の対応は対応を表わすレコード型と2つの親子

集合型で表現しなければならない。図2はレコード型RとSの多対多の対応を表わしたバックマン線図である。レコード型RSはRとSの対応をレコードとするもので、中間(関連)レコード型と呼ばれる。RSはデータを持たなくともよい。レコード型の対応が1対1のものは、親子集合型とするのは冗長で、1つのレコード型として表現できる。

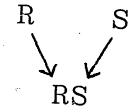
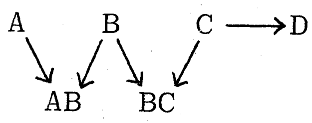


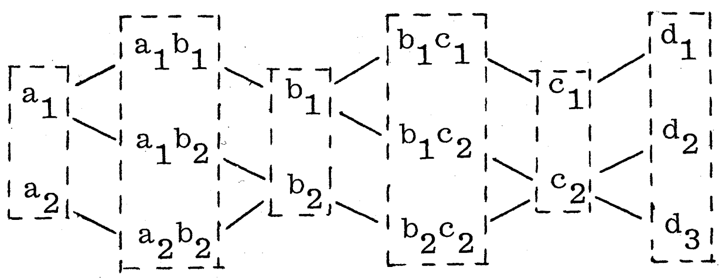

図2

## 3.2 バックマン線図に対応する属性集合

非巡回バックマン線図はMVD集合を表わしている。図3(a)のバックマン線図とそのレコードの親子集合を表わした図3(b)から得られる属性の関係が図3(c)である。Bの値を1つ決定すると対応するAの値の集合とCDの値の集合が




(a)




(b)

| A | B | C | D |
|---|---|---|---|
| $a_1$ | $b_1$ | $c_1$ | $d_1$ |
| $a_1$ | $b_1$ | $c_1$ | $d_2$ |
| $a_1$ | $b_1$ | $c_2$ | $d_3$ |
| $a_1$ | $b_2$ | $c_2$ | $d_3$ |
| $a_2$ | $b_2$ | $c_2$ | $d_3$ |

(c)

A — B — C — D

(d)

図 3

決定し、それらは直積的に対応する。つまり、B→→A|CD のMVDが成り立っている。同様に、C→→AB|D，BC→→A|D のMVDが成立していることが分かる。このように、非巡回バックマン線図はMVD集合を表わしており、次の性質がある。

性質1：非巡回バックマン線図では、その連結部分の属性集合からその部分を除いたバックマン線図の連結部分の属性集合にMVDがある。

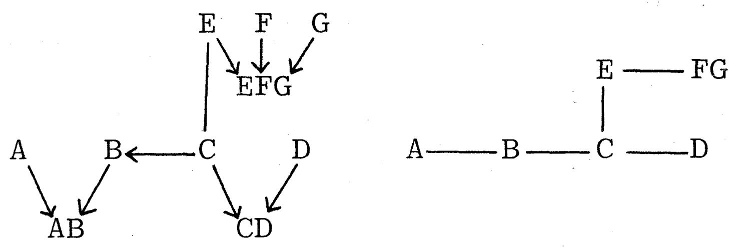
図3（a）で、レコード型 B，BC，C をMVDのキーにとると性質1より BC→→AB|D のMVDが得られるが、これは BC→→A|D と等価である。このようにレコード型の対応のみを表わす中間レコード型は、バックマン線図から得られるMVD集合を求めるのに冗長となるものである。また、MVDは値の対応関係が何対何であるかを問題としない。バックマン線図の中間レコード型を1つの枝で置き換え、親子集合型の矢印を無向としたグラフを新たに導入し、スキーマグラフと呼ぶ。スキーマグラフはバックマン線図の表現するMVD集合を表わしたものである。図3（a）のスキーマグラフは図3（d）となる。

矛盾を起こすMVD集合は、キー分割不整合か推移的不整合のいずれかを起こす[SCI08104]。非巡回スキーマグラフから得られるMVD集合はこの2つの不整合を起こさないの

ので無矛盾である。

性質2：非巡回バックマン線図の表わすMVD集合は無矛盾である。

無矛盾なMVD集合は1つのAJDに等価である結果が得られている[BEERF8105]ので、非巡回バックマン線図は1つのAJDを表わしているとも言える。




(a)　(b)

図　4

スキーマグラフからは、1つの枝で結ばれる節の属性集合の全体に成り立つAJDが求められる。図4(a)のバックマン線図のスキーマグラフは図4(b)、その表わす非冗長なMVD集合は{B→→A|CDEFG，C→→AB|D|EFG，E→→ABCD|FG}、AJDは*[AB,BC,CD,CE,EFG]である。

## 4．関係データベース理論との対応

### 4．1　非巡回バックマン線図と普遍関係仮定

関係データベース理論の多くは普遍関係仮定の上でのもので、任意の属性間の関係は1つの意味しか表わさない。しかし、ネットワークモデルでは複数の意味を持ってよく、この仮定を必要としない。図5(a)は病院と患者の関係を表わす

バックマン線図である。
親子集合型"入院"は、病院と入院している患者の対応を表わしている。レコード型"検査"は患者が受けた検査のデータで、

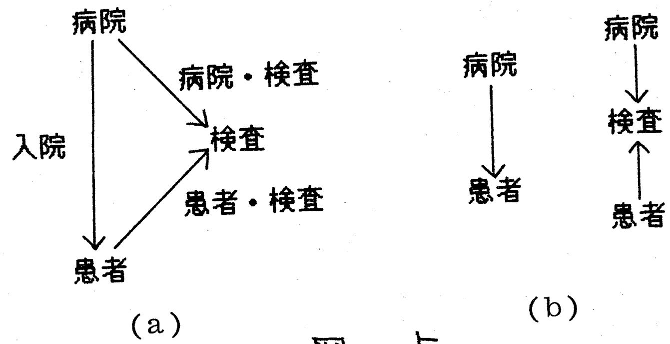



図 5

親子集合型"病院・検査"はその検査がどの病院で行なわれたか、"患者・検査"はその検査はどの患者のものかを表わしている。患者に対応する病院は、入院先の病院と検査を受けた病院の2種類がある。このバックマン線図は普遍関係仮定を満たす2つのバックマン線図(図5(b))を重ねたものになっている。このようにあるレコード型から別のレコード型への経路が複数個あった場合、一般に2つのレコード型の関係は経路の数だけ存在する。逆に普遍関係仮定を満たす関係をバックマン線図で表わしたときに閉路ができるような構成にした場合、2つのレコード型の関係を求めるときに複数の経路が存在するものがあるので、どの経路を通っても同じ関係が得られるように外部からの管理を必要とする。

バックマン線図に閉路ができてもそれを閉路と見なすべきでない場合がある。図6は衝突を起こすMVD集合{AB→→CD|E, AC→→D|BE}をバックマン線図に反映したもので、レ

コード型A,B,Cは、AB,ACの転置ファイル或は索引レコード型とでも言うべきものである。DとEの関係を求める場合、実線の親子集合型をたどって求める経路とレコード型Aを通る経路が存在するが、破線の親子集合型はMVD集合を表わすものではないので、実線の親子集合型を通る経路によらなければならない。関係モデルで衝突を起こすMVD集合が扱えないのは、MVDによって分解した関係の結合が情報損失結合になるものができるためである。この例では、関係は$R_1$(ABE),$R_2$(ABC),$R_3$(ACD)の3つに分解され、$R_1$と$R_3$の結合が情報損失結合となる。これは、バックマン線図でレコード型Aを通る経路に相当している。ネットワークモデルでは、レコード型の関係を求める経路を区別することによって衝突を起こすMVD集合も反映できる。

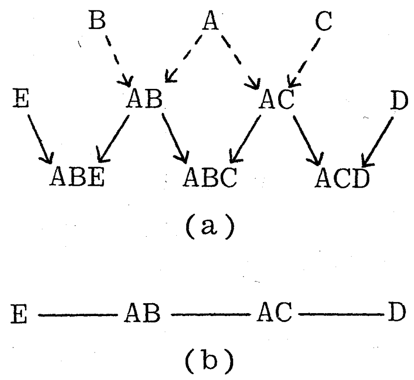
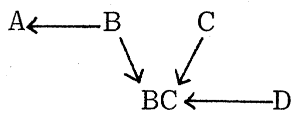



図 6

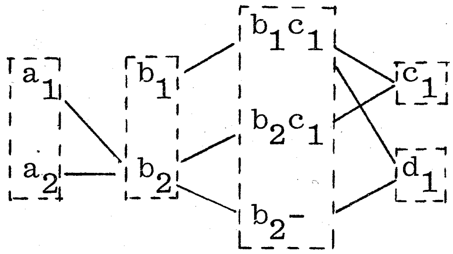
## 4.2 対象集合に関する制約

関係モデルにおける空値は、ネットワークモデルでは子レコードのない親子集合や親子集合に参加しない子レコードにより親子集合型で対応するレコードがないことによる。図7

(a)のバックマン線図とそのレコードの親子集合を表わした(b)から得られる属性A,B,C,Dの関係が(C)である。




(a)




(b)

| A | B | C | D |
|---|---|---|---|
| $a_1$ | $b_2$ | $c_1$ | - |
| $a_1$ | $b_2$ | - | $d_1$ |
| $a_2$ | $b_2$ | $c_1$ | - |
| $a_2$ | $b_2$ | - | $d_1$ |
| - | $b_1$ | $c_1$ | $d_1$ |

(c)


(d)

図 7

親子集合型〈B,A〉で、$b_1$を親レコードとする親子集合の子レコードは存在しないため、$b_1$に対応するAの値は空値となっている。親子集合型〈D,BC〉では、子レコード型BCのレコード$b_2c_1$は親子集合に参加していないので、対応するDの値は空となる。また、BCのレコード$b_2-$のように、部分値のレコードも存在しえる。

これらのことから、2つのレコード型のレコードの関係が存在したとき、バックマン線図或いはスキーマグラフで2つのレコード型を結ぶ経路の途中のレコード型では対応するレコードが必ず存在することになる。図7の例では、AとCの値の関係があったとき、レコード型B,BCで必ず対応するレコードが存在するので、そのA,Cに対応するBの値が必ず存在する。AとDの値の関係があれば、レコード型B,BCの対

応するレコードが存在するが、BCは部分値であってもよいのでA,Dに対応するB又はBとCの値が存在する。

従って、ネットワークモデルで可能な対象は、スキーマグラフの連結部分の各節点から1つ以上の属性を集めたものに、1つの枝で結ばれる2つの節点に共通に現われる属性を加えたものとなっている。逆に、対象がそのネットワークモデルで表現可能かどうかも判定できる。

性質3：対象がネットワークモデルで表現可能であるのは、対象に現われる属性を含むスキーマグラフの節点は連結で、それらの節点からなる部分グラフで対象が隣接節点の共通属性を含むときである。

## 5．ネットワークモデルでの空値を考えた従属性

### 5．1　Lienの研究

ネットワークモデルと関係データベース理論の対応に関する研究はほとんどなく、Lienが1982年のJACMに論文を発表している程度である[LIEN8204]。この論文では、無矛盾で無衝突な空値を考えたMVD（NMVD）の集合は、非巡回バックマン線図で表わされるネットワークモデルと制約的に等価であることが示されている。しかし、この論文には次のような問題点があり、本節ではそのうち（3）について検討

する。

(1) 4.1で述べたように、衝突を起こすMVD集合は非巡回バックマン線図で表現できるにもかかわらず扱い不能としている。

(2) FDについて考察していないので、バックマン線図が冗長なものとなっている。

(3) NMVDの定義が不備であり、この定義では無矛盾・無衝突のNMVD集合と非巡回バックマン線図が制約的に等価であるとは言えない。

Lienは NMVDの定義と同時に空値を考えたFD(NFD)の定義も与えている。これらの定義は次のものである。

NMVD:関係R(U)で、Uの部分集合X,Yが次の性質を満たすとき、XよりYにNMVDがあるといい X→→Y で表わす。

Z=U−X−Yとすると、X全値であるRの任意の組tに対し、(R[XZ=t[XZ]])[Y]=(R[X=t[X]])[Y]

NFD:関係R(U)で、Uの部分集合X,Yが次の性質を満たすとき、XよりYにNFDがあるといい X→Y で表わす。

RのX全値である任意の組r,tに対し、

r[X]=t[X] ならば r[Y]=t[Y]

Lienは このNMVDを用いて、非巡回バックマン線図からNMVD集合を求めるアルゴリズム、NMVD集合から非巡

回バックマン線図を求めるアルゴリズムを示し、ファイルの変換アルゴリズムと合わせて関係データベースとネットワークデータベースの等価性をも示している。

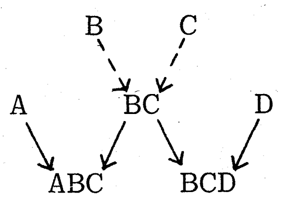
## 5.2 空値を考えた多値従属性：NMVD

非巡回バックマン線図で表わされるネットワークモデルは、制約的にはNMVD集合と等価であるという結果が Lien によって示されている。しかし、Lien のNMVDの定義では空値の扱いが関係モデルとネットワークモデルで異ってしまう。そこでこの定義に2つの修正を加え、ネットワークモデルでも扱えるNMVDを定義する。

図8(a)の関係があったとき、Lien の定義ではNMVD BC→→A|D が成立する。これをバックマン線図、スキーマグラフで表わすと、それぞれ図8(b),(c)となる。このバックマン線図からは、属性A,B,C,Dの値の組み合せで(−0−1)は得られず、本来は存在しなかった(10−0)があらたにできてしまう。BC のレコード型に部分値のものを許さ

| A | B | C | D |
|---|---|---|---|
| 0 | 1 | 1 | − |
| 1 | 1 | 1 | − |
| 0 | 1 | 0 | 1 |
| − | 0 | − | 1 |
| 1 | 0 | − | 0 |

(a)




(b)


(c)

図 8

ないことにすると、BCには(0−)がなくなるため、ABの関係(10)とBDの関係(01)(00)が失なわれてしまう。これは、LienのNMVDの定義ではキーが部分値のものを全く無視しているために起こるものである。この例からも分かるように、ネットワークモデルでNMVDを考える場合、キーが部分値の場合も考慮しなければならない。従って、NMVDの定義はその成り立つための条件を"X全値である任意の組tに対し…"から"Xが空値でない任意の組tに対し…"と修正すべきである。

もう1つの問題点は、キーが全値でないとき(上の修正後はキーが空値のとき)に従属基に対して何の制約も課していないことである。図9(a)の関係は$B \twoheadrightarrow A|C$が成立し、バックマン線図、スキーマグラフでそれぞれ図9(b),(c)と表現できることになる。しかし、このバックマン線図からは属性A,B,Cの値の組み合せ(3−3)を得ることはできない。また、対象としてACがあることから性質3により図9(c)のスキーマグラフでは表現できないことが分かる。このようなことが起きないようにするには、NMVDの定義でNMVDが成り立つ

| A | B | C |
|---|---|---|
| 1 | 1 | − |
| 0 | 1 | − |
| 1 | 0 | 1 |
| − | 2 | 1 |
| − | − | 2 |
| 3 | − | 3 |

(a)

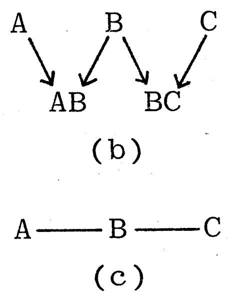



図 9

ための必要条件に"Xが空値のとき、Yが空値でなければZ(=U−X−Y)は空値である"という条件を加えるべきである。

以上の2つの修正から、本稿で新たに導入したNMVDの定義次のようになる。

NMVD：関係R(U)で、Uの部分集合X,Yが次の性質を満たすとき、XよりYにNMVDがあるといい $X \twoheadrightarrow Y$ で表わす。

$Z=U-X-Y$ とすると、Xが空値でないRの任意の組tに対し、$(R[XZ=t[XZ]])[Y]=(R[X=t[X]])[Y]$ かつ $t[X]$ が空値であるとき $t[Y]$ が空値でなければ $t[Z]$ は空値である。

5.3 空値を考えた関数従属性：NFD

NMVDの定義の修正と同様に、NFDの定義もネットワークモデルの構造に適合するように修正した方がよい。図10(a)の関係は、Lienの定義によるNFD $AB \to C$ を満たしている。しかし、ABの値(0−)に対してCの値1,0の2つが存在するのは不自然であり、ネットワークモデルでも図10(b)の親子集合型で表現することはできない。NFDを親子集合型と対応させ、NFDのキーが部分値のときも右辺の値

| A | B | C |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | − | 1 |
| 0 | − | 1 |
| 0 | − | 0 |

(a)


(b)

図 10

が１つ定まるように、ＮＦＤの定義を次のように修正する。

<u>ＮＦＤ</u>：関係$R(U)$で、$U$の部分集合$X,Y$が次の性質を満たすとき、$X$より$Y$にＮＦＤがあるといい $X \rightarrow Y$ で表わす。

Rの任意の組$r,t$に対し、$r[X],t[X]$が空値でないとき

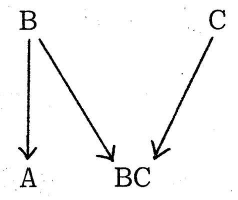
$r[X]=t[X]$ ならば $r[Y]=t[Y]$

ネットワークモデルでの従属性を考えた場合、ＮＭＶＤの第１の修正やＮＦＤの修正が必要となるのは、レコード型に部分値のレコードが存在しえるためである。これらの修正により、ネットワークモデルに多値従属性や関数従属性を空値を含めて適用できるが、このために推論則"$X \rightarrow Y$ ならば $X \twoheadrightarrow Y$"は成立しない。図11(a)の関係はこの例である。この関係で成り立つＮＭＶＤ，ＦＤの集合は$\{B \twoheadrightarrow A|C,\ A \rightarrow B\}$であり、$A \twoheadrightarrow B$ は成立しない。従ってこの関係をバックマン線図に表わしたものは図11(b)となる。これはＮＭＶＤの第２の修正により、$X$が空値のときの$Y$と$Z=U-X-Y$の値の存在に制限をつけたために起こる。

このことから、ＮＦＤ集合$\{X \rightarrow Y,\ Y \rightarrow Z\}$は図12のネットワークモデルで必ずしも表現

| A | B | C |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 0 | 1 |
| 1 | 1 | – |
| 2 | 1 | – |
| 3 | 2 | 2 |
| – | 3 | 2 |

(a)




(b)

図 11

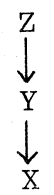



図12

できるとは限らない。これは、Xのある値に対してYの値が空でZの値があり、その組がX→Zに反している場合があるためである。この表現が可能となるのは、NMVD：Y↠X｜Zが成り立つときで、このときはNFD：X→Zも成り立つ。この例からも分かるように、ネットワークモデルの構造はNMVD集合によるもので、NFDは対応関係の特別な場合として表現できるものにすぎない。

## 6．ネットワークモデルに対するスキーマ設計

4節で述べたように、普遍関係仮定のもとではネットワークモデルの基本構造は非巡回スキーマグラフで表現できる。スキーマグラフは無矛盾なNMVD集合、つまりは1つのAJDを表わしたものであるから、ネットワークモデルに対するスキーマは、

(1) すべての属性を普遍関係仮定が成立する部分に分ける。
　　（属性は複数の部分に属してもよい）
(2) 分類された各部分で成り立つNMVD集合(AJD)を求める。
(3) 各部分に対しNMVD集合(AJD)からスキーマグラフを作る。
(4) 全てのスキーマグラフを重ね1つのスキーマグラフにする。
(5) 対応関係により中間レコード型を追加し、成り立つNFD集合から親子集合型を決定する。

(6) スキーマを最適化する。

の手順によって設計できる。

NMVD集合(AJD)からスキーマグラフを作成する方法は、NMVD集合のキーを用いる方法とAJDを表わすハイパーグラフを用いる方法がある。非冗長NMVD集合$\{X_i \twoheadrightarrow Y_{i1}|Y_{i2}|\cdots|Y_{iim} : i=1,2,\cdots,n\}$から得られるスキーマグラフの節点集合は$\{X_i, Y_{1j_1} \cap Y_{2j_2} \cap \cdots \cap Y_{nj_n} : i=1,2,\cdots,n, j_k=1,2,\cdots,k_m\}$である。この節点集合と、$X_i$に含まれる属性を除いたときに連結となる部分の属性集合が$Y_{i1}, Y_{i2}, \cdots, Y_{iim}$に一致するように枝集合を決定したものがスキーマグラとなる。ハイパーグラフでは、すべての関節点集合に対してそれに含まれる属性集合と、それらをすべて除いたときの連結部分の属性集合がスキーマグラフの各節点に対応する。ハイパーグラフの関節集合を除いたときの連結部分の属性集合が、その関節集合の属性をスキーマグラフからすべて除いたときの連結部分の属性集合と一致するようスキーマグラフの枝集合を決定する。

7. あとがき

空値の扱いは関係データベース理論でも完全には明確にされておらず、空値を非存在と不明に分けて考えた場合の扱いの違い、空値を1つの値として見たときの意味といった問題

がある。ネットワークモデルでこれらの問題をどう扱うかを考えると同時に、関係モデルにおける扱いも考えなければならない。修正後のNMVD,NFDについても、多少その性質について述べたが、推論則を明確にし、健全で完全な公理系を求める必要がある。本稿ではネットワークモデルの表現できる意味制約について基本的な考察を行なった。この結果を用いて、さらにネットワークモデルに対するスキーマ設計手法へと発展させる予定である。

謝辞　御討論いただいた矢島研究室の諸氏に感謝いたします。

## 参考文献

[AHO-B7970] Aho,A.V., Beeri,C., and Ullman,J.D., "The Theory of Joins in Relational Database", ACM TODS, Vol.4, No.3, PP.297-314, Sept. 1979.

[BEERB7809] Beeri,C., Bernstein,P.A., and Goodman,N., "A Sophisticate's Introduction to Data Base Normalization Theory", Proc. of 4th International Conferrence on VLDB, pp.113-124, Sept. 1978.

[BEERF8105] Beeri,C., Fagin,R., Maier,D., Mendelzon,A., Ullman,J., and Yannakakis,M., "Properties of Acyclic Database Schemes", ACM SIGACT Symposium on the Theory of Computing, pp.355-362, May 1981.

[DATE81] Date,C.J., An Introduction to Database Systems, 3rd Edition, Addison-Wesley, 1981.

[FAGI7709] Fagin,R., "Multivalued Dependencies and New Normal Form for Relational Database", ACM TODS, Vol.2, No.3, pp.262-278, Sept. 1977.

[GOLD8109] Goldstein,B.S., "Constraints on Null Values in Relational Database", Proc. of 7th International Conferrence on VLDB, pp.101-110, Sept. 1981.

[KAMB82] 上林弥彦, "データベースの基礎理論(1)(2)(3)", 情報処理学会誌, 第22巻, 第3,7,9号, 1982.

[LIEN7910] Lien,Y.E., "Multivalued Dependencies with Null Values in Relational Databases", Proc. of 5th International Conferrence on VLDB, pp.61-66, Oct. 1979

[LIEN8204] Lien,Y.E., "On the Equivalence of Database Models", JACM, Vol.29, No.2, pp.333-362, April 1982.

[SCIO8104] Sciore,E., "Real World MVD's", Proc. of ACM SIGMOD International Conference on Management of Data, pp.121-132, April 1981.

[TSICL82] Tsichritzis,D.C., Lochovsky,F.H., "DATA MODELS", Prentice-Hall software series, 1982.

[ZANI7607] Zaniolo,C., "Analysis and Design of Relational Schemata for Database Systems", Computer Methodology Group Report, UCLA-ENG-7669, Dept. of Computer Science, University of California at Los Angels, July 1976.