数理解析研究所講究録
461巻 1982年 173-186

関係データモデルへの限量の導入について

国際情報研　竹島 卓

1. まえがき

データベースにおいて実際のデータをとり扱う際には、ある情報が未知であるとか、もともと存在し得ない情報であるというような状況がよく起こる。この種の情報を陽にデータベースで取扱うために、いわゆるnull値が導入されている。一般にnull値によって表現されることを意図されている情報は、プラグマチックな要求から分析すると、優に10種を越えるといわれている。[1]　これらのうちで、比較的なじみやすく、かつ利用度も多い2種のnull値、存在null値と非存在null値、を最初に論じる。また、これらとは違った使われ方をするnull値（無情報）についても簡単にふれる。

2 不完全情報をもつデータベース

Vassiliou[2], Lipski[3], 古川[4]らによれば、不完全情

報をデータモデルの中に含ませることによって，上述の２種のnull値，存在null値と非存在null値とが不完全情報の両極端、無情報と過剰情報として取扱えるとした．特に古川，Lipskiは，一般にOR-factといわれている確定的でない情報をデータベースで扱い，質問応答において適切な答を得る方法について論じた．この要点は，データ項がすべて（基本的には）集合値である場合を扱うことにあり，その集合は，本来の関係でそのデータ項目位置に来るはずのアトミックなデータの動き得る範囲をあらわしている．singleton集合はその唯一の要素と同一視(?)すれば，従来の関係モデルのあらわしている意味があらわせる．特に集合が空集合の場合はデータの動き得る範囲がないこと，つまり矛盾をあらわしている．また，集合が，そのデータ項目位置のデータが走り得る全領域として与えられたときには，そのデータ項目自体は何ら情報を含んでいないことになる．これを未知情報と解釈しようというアイデアである．Vassiliouは，通常のデータ領域に$\top$（top，矛盾）と$\perp$（bottom，undefined）とを追加し，Dennotational Semantics流のデータ意味論を構成した．

これはLipski，古川らでは空集合と全集合とに対応する．

この他にCodd[5]が３値論理よる関係モデルの拡張を論じている．

これらの方法は、いずれもデータ領域(真理値も含めて)を拡張することによって、不確定情報を得ようとするものであるが、真の意味の否定情報はあらわし得ない。Vassiliou, Lipski は矛盾情報によって否定をあらわそうとしているが、データベースから矛盾が導かれることと、否定的事実が導かれることとは全く異なる概念である。運用でカバーできるとはいっても大変不安である。

本稿では、OR-fact, repeating group (AND fact とでもいうべきか) 存在null値, 非存在null値(否定的事実)を統一的に扱うための考え方を提案する。

3 表とその解釈

筆者は、関係モデルが Codd のいう第一正規形をみたしている限りは、関係とは数学的対象であると考えてさしつかえないと考えているが、null値の導入や不完全情報(をあらわすデータ)の導入に際しては、データ領域(意味領域)を複雑にするよりは、データベースに格納される「関係」や「データ項目」は言語の要素と考え、その言語の解釈を単純な意味領域上で与えるべきだと考えている。その理由は、関係と述語との(ある意味での)1対1対応と通常の1階論理の意味論を放棄するのが得策でないと考えるからである。簡単な例

をあげれば，タップル<a，⊥>が関係Rに存在するとき，述語 $P_R(a,\perp)$ の真理値をいくら複雑にしてみても，$\exists y P_R(a,y)$ に相当する意味を与えることは困難である．

ここでは，真に数学的な意味での関係は意味領域にあるものとし，データベースで表面に出るものは，そのような関係を指示する言語要素であると考える．その要素をここでは「表」と呼ぶことにする．表にはつぎの記号列が記入できる．

(i) 通常のアトミックな値（をあらわす記号）．

(ii) 存在null記号　　ε　と

非存在null記号　　ω　．

(iii) 限量記号　*some*, *every*, *no* のうちのいずれかひとつの後に単項関係名が付いたもの．

もちろん，表自身の名前も含まれる：　上記(i)～(iii)を「表項」と呼ぶ．表項の n 組を関係データモデルに模してタプルと呼ぶ．タプル $\mathbf{r}=\langle r_1, r_2, \cdots, r_n\rangle$ が表 $\mathbf{R}$ に含まれることは、$\mathbf{r}\in\mathbf{R}$ と書くことにする．

従来の関係の見方のひとつによれば，関係達は，述語論理の文集合（セマンチック・コンストレイントやキュエリー等）に対する解釈を定めるひとつのモデルを構成している．また別の見方をすると、証明論的な *Axiom* の集合とも見なせる．後者の見方は知識ベース応用でしばしば採用される方式で，

関係を数学的な意味で（モデルとして）とらえるのではなく言語ととらえる見方である．そこでは，関係は同じ情報をあらわす述語論理の文集合へと翻訳される．翻訳はつぎの式(1)に従う．

データベースに**R**という関係があってタプル**r**が**r**∈**R**となっている．

$\Longleftrightarrow$ $R(r_1, r_2, \cdots, r_m)$ が Axiom として与えられる．　(1)

ところが，アトミックとはみなせない値（をあらわす記号）を表中に記述することを許すと，(1)の形の翻訳は適用できない．しかしながら，Montague 文法で用いるような高階の汎関数を用いれば，つぎのように翻訳できる．

データベースに**R**という表があってタプル**r**が**r**∈**R**となっている．

$\Longleftrightarrow$ $r_1^*(\lambda x_1.\, r_2^*(\lambda x_2. \cdots r_m^*(\lambda x_m.\, R(x_1, x_2, \cdots, x_m)) \cdots)$　(2)

が Axiom として与えられる．

ここに各 r_i^* は r_i の翻訳でつぎの翻訳規則に従うものとする．

(i°) アトミックな値 a に対して $a^* \triangleq \lambda P.\, P(a)$,

(ii°) $\varepsilon^* \triangleq \lambda P.\, \exists x P(x)$, $\omega^* \triangleq \lambda P.\, \forall x \sim P(x)$,

(iii°) $(\text{some } S)^* \triangleq \lambda P.\, \exists x (S(x) \wedge P(x))$,

$$(\text{every } S)^* \triangleq \lambda P.\ \forall x\,(S(x) \to P(x)),$$

$$(\text{no } S)^* \triangleq \lambda P.\ \forall x\,(S(x) \to \sim P(x)).$$

この翻訳規則はアトミックな値に対しては従来の解釈に一致している.

$A_1, A_2, \cdots, A_n$ が汎関数で, $P[x_1, x_2, \cdots, x_n]$ が変数 $x_1, x_2, \cdots, x_n$ を自由に含む式とするとき, 簡単のためつぎの記法を用いる。

$$\left|\begin{matrix} A_1 & A_2 & \cdots & A_n \\ x_{\sigma(1)} & x_{\sigma(2)} & \cdots & x_{\sigma(n)} \end{matrix}\right| P[x_1, x_2, \cdots, x_n] \tag{3}$$

$$\triangleq A_1(\lambda x_{\sigma(1)}.\ A_2(\lambda x_{\sigma(2)}. \cdots A_n(\lambda x_{\sigma(n)}.\ P[x_1, x_2, \cdots, x_n]) \cdots))$$

ここに σ は $1, 2, \cdots, n$ の置換である. この記法によれば翻訳規則式(2)は,

$$\left|\begin{matrix} r_1^* & r_2^* & \cdots & r_n^* \\ x_1 & x_2 & \cdots & x_n \end{matrix}\right| R(x_1, x_2, \cdots, x_n) \tag{4}$$

と簡明に書ける.

さて, このように提案する表の解釈では解決できない問題がひとつある. それは自然言語にもある多義性の問題である.

$1, 2, \cdots, n$ の任意の置換 σ に対して

$$\left|\begin{matrix} r_1^* & r_2^* & \cdots & r_n^* \\ x_1 & x_2 & \cdots & x_n \end{matrix}\right| P[x_1, x_2, \cdots, x_n] = \left|\begin{matrix} r_{\sigma(1)}^* & r_{\sigma(2)}^* & \cdots & r_{\sigma(n)}^* \\ x_1 & x_2 & \cdots & x_n \end{matrix}\right| P[x_1, x_2, \cdots, x_n] \tag{5}$$

が成立つとき，タプル $r=\langle r_1, r_2, \ldots, r_n \rangle$ は一義的であるといい
そうでないときは多義的であるという。 多義性は限量および否定の順序が入れ替ることによって生じ，この回避のためにはタプルの構成順序を明示する必要がある。 ここでは，一義的な場合のみを扱う。 すなわち，1タプルにおいてはつぎの制限された限量しか扱わない。

(i) ε と *some* とは混在と重複出現が可能。

(ii) *every* は重複出現が可能。εあるいは*some*と混在不可。

(iii) *no* もしくは ω は単独で出現し，重複や他の限量との混在は許されない。

4 <u>例題</u>

xがyの子供であるという関係をCHLD(x,y)であらわす。 これに対応する表が(a)に示されている。 またSとTというふたつの単項関係がそれぞれ表(b)と表(c)とで与えられている。

CHLD

X	Y
Mary	John
ε	Tom
ω	Jack
some **S**	Bill
every **T**	Jim

(a)

S

U
Joe
Ann

(b)

T

V
Bess
Ed

(c)

<u>(a)の第1タプルから</u>

$$\left|\begin{matrix} Mary^* & John^* \\ x & y \end{matrix}\right| CHLD(x,y) = Mary^*(\lambda x.\ John^*(\lambda y.\ CHLD(x,y)))$$

$$= \quad Mary^*(\lambda x.(\lambda P.P(John))(\lambda y.CHLD(x,y)))$$

$$= \quad Mary^*(\lambda x.(\lambda y.CHLD(x,y))(John))$$

$$= \quad Mary^*(\lambda x.CHLD(x,John))$$

以下 同様に

$$= \quad CHLD(Mary, John) \,. \tag{6}$$

これは、通常のタプルの解釈と一致する.

(a)の第2タプルから

$$\left|\begin{matrix} \varepsilon^* & Tom^* \\ x & y \end{matrix}\right| CHLD(x,y) = \exists x\ CHLD(x, Tom) \,. \tag{7}$$

これは、Tom の子供の存在を主張しており, ここから ε を存在null値と呼ぶ.

(a)の第3タプルから

$$\left|\begin{matrix} \omega^* & Jack^* \\ x & y \end{matrix}\right| CHLD(x,y) = \forall x \sim CHLD(x, Jack) \,. \tag{8}$$

これは、Jack の子供の不存在を主張しており、ここから ω を非存在null値と呼ぶ。 このこと自体は矛盾ではないことに注意して欲しい. Lipski の φ(空集合) や Vassiliou の **T** は矛盾であるので、それでは不存在情報はあらわせない.

(a)の第4タプルから

$$\left|\begin{array}{cc}(\text{some } S)^* & Bill^* \\ x & y\end{array}\right| CHLD(x,y) = (\text{some } S)^* \lambda x.\ Bill^*(\lambda y.\ CHLD(x,y))$$

$$= (\text{some } S)^* \lambda x.\ CHLD(x, Bill)$$

$$= (\lambda P.\exists z(S(z) \wedge P(z)))\ \lambda x.\ CHLD(x, Bill)$$

$$= \exists z(S(z) \wedge CHLD(z, Bill)) \tag{9}$$

ここで S に関して閉世界仮説を導入することにより

$$\forall z(S(z) \to (X=Joe) \vee (X=Ann)) \tag{10}$$

が得られる。(9)と(10)との帰結として、

$$CHLD(Joe, Bill) \vee CHLD(Ann, Bill) \tag{11}$$

が得られる。これは OR-fact をあらわしている。

(b)の第5タプルから

$$\left|\begin{array}{cc}(\text{every } T)^* & Jim^* \\ x & y\end{array}\right| CHLD(x,y)$$

$$= \forall z(T(z) \to CHLD(z, Jim)) \tag{12}$$

が得られる。一方 T からは

$$T(Bess)\ ,\ T(Ed) \tag{13}$$

が閉世界仮説とは関係なしに得られるので、(12)と(13)との帰結として

$$CHLD(Bess, Jim)\ ,\ CHLD(Ed, Jim) \tag{14}$$

が得られる。これは repeating group をあらわしている。

以上の例に加えてつぎのことを注意しておく．

Sに対する閉世界仮説の下では，**S** = 空集合とするとそれから矛盾が生じる．すなわちデータベースの矛盾であり，LipskiのØ，Vassiliouの**T** に相当する．

一方，**T**に対する閉世界仮説の下では，**T** = 空集合は，Jimに子供がいないことをあらわしている．すなわち否定的事実の表明である．

5．関係の射影と無情報null値

この節では，前節までに定義した2種のnull値とは異なる第3のnull値，Λであらわす，について簡単に述べる．これは本講究録中の田中譲氏のnull値⊥と（おそらく）同じである．田中氏は，⊥を含む関係の分解合成を，⊥を考慮したnatural dependency に従って行えばよいことを示した．ここでは，⊥と関連して，関係の射影の意味を考えなおしてみたい．実は射影ばかりでなく，一般に関係演算の結果についても，その意味を考えなおしてみる必要がある．（特に関係の外延性，内包性についても）．

簡単のために3項関係を例として考察する．属性をA, B, Cとしよう．その表を**R**とする．表**R**の意味としてその述語翻訳 Rがあるとする．**R**がごくふつうの関係ならば，

その AB-射影 $\boldsymbol{R}[AB]$ の述語的意味を R_{AB} とすると，

$$\begin{aligned} &\langle a, b\rangle \in \boldsymbol{R}[AB] \\ &\quad\Longleftrightarrow \quad R_{AB}(a, b) \end{aligned} \tag{15}$$

であって，

$$R_{AB}(a, b) = \exists z\, R(a, b, z) \tag{16}$$

が成立する． すなわち，通常の関係の射影の意味は，式(16)で定められる． これは射影でのタプル $\langle a, b\rangle$ はもとの関係では存在null値 ε を含むタプル $\langle a, b, \varepsilon\rangle$ に相当することを含意している．

しかしながら一方では，これと異る取扱いのnull値を持ちこまねば解釈できない奇妙な関係モデルの取扱いが通用している． いわゆる従属性による分解合成の議論は，ユニバーサルリレーションの存在仮定と，属性の意味の唯一性仮定との下で，updating の anomaly を解消することを目的としていたはずであるが，現実には，分解可能性の理論は基本的には関係の外延（オカレンス，インスタンともいう）に対してのみ適用可能であって，内包に対する操作であるupdatingや，分解し，合成するというオペレーションに対して単純に適用可能とはいえない．

例として属性ABCをもつリレーション$\boldsymbol{R}$を考える． 従属性理論に従えば，$\boldsymbol{R}$に関数従属 $A \to B$ があれば，$\boldsymbol{R}$は射影，

$R[AB]$と$R[AC]$とに分解できる．すなわち

$$R = R[AB] \bowtie R[AC] \qquad (17)$$

が成立する．ここで $\langle a,b,c\rangle \in R \Leftrightarrow$ 'aの年令はb，かつcはaの子供'という意味であるものとしよう．すると，特にaに子供が一人もいない場合には，この情報はRではうまくあらわせない．(むしろaの年令がbという情報)

$$R(a,b,c) = \text{'}a\text{の年令は}b\text{，かつ}c\text{は}a\text{の子供'}$$
$$= Age(a,b) \wedge child(a,c) \qquad (18)$$

とすると，子供がいないことをあらわすために，ωを用いるとしてもうまくいかないのである．すなわち，

$$\langle a,b,\omega\rangle \in R$$
$$\Leftrightarrow \quad \forall c \sim R(a,bc)$$
$$= \quad \forall c\,(\sim Age(a,b) \vee \sim chid(a,c)) \qquad (19)$$

では，$\sim Age(a,b)$ かあるいは $\sim\forall c\ child(a,c)$としか結論できないし，むしろ言いたかった肯定的事実$Age(a,b)$の情報は消失してしまう．だからこそ，分解しておくのだということには心情的には賛成である．しかしそれを正当化するには，(17)式の成立を放棄する~~必要~~がある．つまり今までの従属性理論で扱っていたような数学的なきれいごとの関係での(17)式，その意味である(18)との対応といったことについて根本的に考え直さねばならない．

解決の方針は、関係をtotal（全域的）なものでなく、partial（部分的）なものでよいと譲歩することと、その未定義値の解釈のために、関係に対してその解釈ベースとしての複数の述語（射影に関するmeaning postulateは必要）を割当てること（今までは単一の述語でよかった。），およびその意味に従う、射影，結合などの関係上の演算を再定義することである。

partialな関係を考え、その上で射影結合を再定義する点は、田中譲氏の方針と（おそらく）一致している。筆者の興味は、述語的に解釈できる意味をどのように構成できるかという点にある。この方向での成果については今回は割愛する。

6 あとがき

本稿では、関係データモデルについての、理論と現実の運用との間のギャップを埋めるためにとるべき2つの方法を示した。ひとつは、関係モデルはそのままにしておいて、現実におけるデータベースの表現言語としての表を考えることであり、今ひとつは、関係モデルの関係そのものを全域的なものから部分的なものへと一般化することである。このこと関連する概念として、null値と呼ばれている、特別な値(?)

の意味について，εとωとの２種類を限量の導入によって定めた． またこれらとは異るnull値（部分関数値としての未定義値）の必要性について，ひとつの例をあげて説明した．未定義値が必要となる理由はこの例にとどまらずいくつかのものがある。 関係の部分化については十分述められなかったが，機会をあらためて詳述したい．

謝辞　関係についての見直しについてカテゴリー論的見地から，多くの有益な知識や定式化のアイデアを提供してくれた，国際研の加藤研究員に感謝します．

文献

[1] CODASYL Develop. Comittee, An Information Algebra, Phase I Report, CACM, 1962.

[2] Vassiliou, Y.: Null Values in Data Base Management, Proc. ACM SIGMOD, 1979

[3] Lopski, W. Jr., On Semantic Issues Connected with Incomplete DataBase, VLDB, 1977

[4] 古川康一，評価法の拡張による未知情報の扱いについて，情処研メモ

[5] Codd, E.F., Extending the Database Relational Model to Capture More Meaning, ACM TODS, 1978