数理解析研究所講究録
461巻 1982年 265-277

タプル重みづけとしきい値を導入した関係演算とその応用

電電公社武蔵野研究所　　斉藤宗昭

1，　まえがき

関係データベースは，その基本演算が体系的に確立しているため，それを前提とする意味表現，検索言語，演算マシンの研究が盛んである。　しかし関係演算マシンの適用拡大を考える立場から，集合演算の枠の中で演算自体の拡張を考える余地もある。　その拡張には2種類考えられる。　例えば部分的な従属関係を切出した更新用ビューを与えて，分解されてない親関係表を全体的に矛盾なく更新する演算は従来の演算を組合せたマクロ演算として実現出来る。　このような演算は演算マシンの高速化とスケジューリングの問題に近い。　より基本的な拡張としては，除表で与えられた条件の満足度が不十分でも商表を出力する関係割算が挙げられる。　この演算を一般化すると従来の重みづけ定量検索を包含しFuzzy 集合検索となる。　このような拡張形割算の実現法として導入したタプル重みづけとしきい値選択を，関

係演算全体に拡大適用することで一つの集合演算体系を得た。これにより，関係表を属性欄数に等しい次元数のマトリクスと対応させて，関係演算によるマトリクスの線形演算が行えるようになり，あわせてしきい値論理による論理積・論理和アレイが組めるようになった。 この応用として，知識工学で使われる確率的推論を関係表と集合演算で実行することが先ず考えられる。

2、 拡張形関係割算

先ず，重みづけとしきい値の導入に到った拡張形関係割算について述べる。 関係表A（氏名，習得外国語）とB（必要外国語）があり，例えばスイスへ出向する適任者を知るために［独語，仏語，伊語］の3タプルからなる除表Bを与えてAを割算し，商表に当該3ヶ国語が出来る人を得るのが従来の割算である。 それを次のように拡張して実用性を向上させることを考える。

（1）例えば除表に［独，仏，伊，英］の4国語4タプルを与え，そのうち3国語が出来る人は商表に出力する（しきい係数3/4）閾値選択機能

（2）スイス出向の場合，独，仏，伊，英語の順に必要性が高いと云う重みづけ（例えば4：3：2：1）機能

（3）各人の外国語のスキルを数値で与えておくと，スキルと

必要度の総合評価で商表に適任者を出力する機能

(4) 必要度に対する否定形の導入，例えば中国語が出来る人は別の任務が多いので原則として欧米出向しないとすれば除表Bの中の中国語にマイナスの重みづけをする機能

(5) 出向先毎に個別に除表を与えるのでなく全出向先を一括して除表B（出向先，必要外国語）に与える演算機能

以上の機能を実現する前提として，スキル，必要度，評価点等を格納する重み係数欄を，属性欄とは別に全ての関係表に持たせることにする。 もし重み係数を持たない関係表があればそのデイフオルト値を1として演算を進める。

3. 拡張形割算の定義

上述の(1)～(5)を包含した割算の定義を手順形で述べ，その定義に必要な演算要素の説明をあとから述べる。例えば関係表A（スキル，氏名，外国語）を除表B（必要度，外国語，出向先）で割算して商表E（評価点，氏名，出向先）を得る拡張形割算である。 これらの表を一般的に $A(a_{ij}, i, j)$，$B(b_{jk}, j, k)$，$E(e_{ik}, i, k)$ と記述する。 なおこの例では属性欄数が2であるが一般には任意である。 また以下の定義では中間結果表C，Dとしきい係数 h を用いている。

① $A(a_{ij}, i, j) \underset{j}{\bowtie} B(b_{jk}, j, k) \Rightarrow C(c_{ijk}, i, j, k)$

ただし　$c_{ijk} \Leftarrow a_{ij}\, b_{jk}$

② $\underset{ik}{\Pi}\, C(c_{ijk}, i, j, k) \Rightarrow D(d_{ik}, i, k)$

ただし　$d_{ik} \Leftarrow \sum_{j} c_{ijk}$

③ $\underset{k}{\Pi'}\, B(b_{jk}, j, k) \Rightarrow W(w_k, k)$

ただし　$w_k \Leftarrow \sum_{j} |b_{jk}|$　（絶対値をとる）

④ $D(d_{ik}, i, k) \underset{k}{\text{Ⓗ}} h\, W(w_k, k) \Rightarrow E(e_{ik}, i, k)$

すなわち　$\begin{cases} d_{ik} \geqq h w_k \text{ なら } e_{ik} \Leftarrow d_{ik} \\ d_{ik} < h w_k \text{ なら } e_{ik} \Leftarrow 0 \end{cases}$

ここでは演算要素記号として $\bowtie$，Π，Ⓗ，Π' を用いており，説明は次章で行う。 もしここで属性欄及び添字の k を削除し，a_{ij}，b_j，h の値を全て1とすると従来の割算の定義と等価になり，最初の例では e_i に3（3国語が出来る）が得られる。 もし商表で得られた候補者に関する内訳理由を関係表 G で得たければ以下の手順を加える。

⑤ $E(e_{ik}, i, k) \Rightarrow F(1, i, k)$

つまり　$f_{ik} \Leftarrow 1$　（初期設定する）

⑥ $C(c_{ijk}, i, j, k) \underset{ik}{\bowtie} F(1, i, k) \Rightarrow G(g_{ijk}, i, j, k)$

なお　$g_{ijk} \Leftarrow c_{ijk}$

ここで属性欄は i：氏名， j：外国語， k：出向先である。

4. 演算要素記号の説明

4.1 自然結合（$\bowtie$）

例えば $\underset{j}{\bowtie}$ は属性 j の実現値が等しいタプルの結合で，両関係表の属性欄集合の包含関係によって内積，外積，直積（対応する属性欄がない）をとり，新しい重み係数欄には両関係表の当該結合タプルが持つ重み係数の積（スカラー量）を与える。

4.2 射影（Π）

例えば $\underset{ik}{\Pi}$ は属性 i，k への射影であるがそれ以外の属性欄を消去するに際して発生する重複タプルはその重み係数を集計して1個のタプルに集約して新しい重み係数とする。また $\underset{0}{\Pi}$ は空属性への射影であり，これを行うと関係表は1個のスカラー量（重み係数の合計値又はタプル数）になる。

4.3 絶対値射影（Π'）

本演算はしきい係数を正規化して与えられるよう可能最大しきい値を計算するための演算で，重複タプルに対する重み係数の集計をその絶対値をとって行う。 負の重み係数は否定形と正の重み係数の複合に過ぎず2重否定を含めた最大可能しきい値を与えるため絶対値をとる。

4.4 閾値選択（Ⓗ）

例えば $\underset{k}{Ⓗ}$ は属性欄 k の実現値が一致するタプル相互間

で重み係数を比較し，しきい値以上であればそのまま出力し しきい値未満であればそのタプルは出力から削除する。 なお重み係数値0は，そのタプルが物理的にも存在しないことと等価である。 なお本演算記号の右側の閾値関係表は演算キー以外の属性欄(演算キーの実現値に関する重複タプル)を持たない。 もしそれに反する時は絶対値射影が自動的に行われるようインプリメントする。

5， マトリックス及びその線形演算との対応

マトリックスとの対応づけに於いては，重み係数がその元素，属性欄数がその次元数，属性欄の実現値がそれと対応する添字の実現値となる。 すなわち添字は連続自然数をとらず任意の文字列のままでそれと同じ役割を果す。 また属性欄が1個しかなければベクトルに対応する。

例えばここに相関確率マトリックス $A=[a_{ij}]$ と所見ベクトル $B=\{b_j\}$ があり，疾患ベクトル C が B の線形変換 $C \Leftarrow AB$ で求められるとすると，この演算は以下の関係演算と等価である。

$$C(c_i, i) \Leftarrow \prod_i \left(A(a_{ij}, i, j) \underset{j}{\bowtie} B(b_j, j) \right)$$

つまり属性欄 i，j に人間がそのまま読み取れる所見や疾患を表示したままでマトリックス演算にのる便利さがあり，しかも属性欄やタプルの配列は見易さ等の都合でどう変えても

演算上は等価である。 関係表ではその次元数に関係なく全元素を縦1列に配置するのでマトリックス表示に比してコンパクトではないが，値が0の元素はタプルが削除される点で逆にコンパクトである。

次の例として，相関確率関係表A(a_{ij}，所見，疾患)をF(f_{il}，所見，病理)とG(g_{lj}，病理，疾患)に分解し，より直交したベクトルである病理をキーファクタとして維持更新することを想定する。 するとFとGからAを得る演算はマトリックスの積 $A \Leftarrow FG$ に対応する関係演算

$$A(a_{ij}, i, j) \Leftarrow \prod_{ij}\left(F(f_{il}, i, l) \underset{l}{\bowtie} G(g_{lj}, l, j)\right)$$

により行える。

6． AND／ORゲートアレイとの対応

前述の拡張形割算により，しきい値を用いたANDを実行した後，射影によりORを実行する。 例えば疾患関係表G(評点，疾患，項内番号，積項番号) = G(C_{ijk}，i，j，k)が与えられたとする。 ここで項内番号，積項番号は論理代数式の例を用いて実現値の例を示すと以下の通りとなる。

項内番号(j)	1 2 3	1 2	1 2 3 4
積項番号(k)	1 1 1	2 2	3 3 3 3
論理代数式	F G H +	I J +	K L M N

関係表 $A(a_{ij}, i, j)$ マトリックス $A=[a_{ij}]$

i：行 j：列

重み係数欄	属性欄	
a_{ij}	i	j
$a_{イ\alpha}$	イ	α
$a_{ロ\beta}$	ロ	β
$a_{ハ\gamma}$	ハ	γ
$a_{イ\gamma}$	イ	γ

実現値

$$
\begin{array}{cc}
 & \overbrace{\begin{array}{ccc} \alpha & \beta & \gamma \end{array}}^{j} \\
i\left\{\begin{array}{c} イ \\ ロ \\ ハ \end{array}\right. & \begin{bmatrix} a_{イ\alpha} & 0 & a_{イ\gamma} \\ 0 & a_{ロ\beta} & 0 \\ 0 & 0 & a_{ハ\gamma} \end{bmatrix}
\end{array}
$$

図1． 関係表とマトリックスの対応例

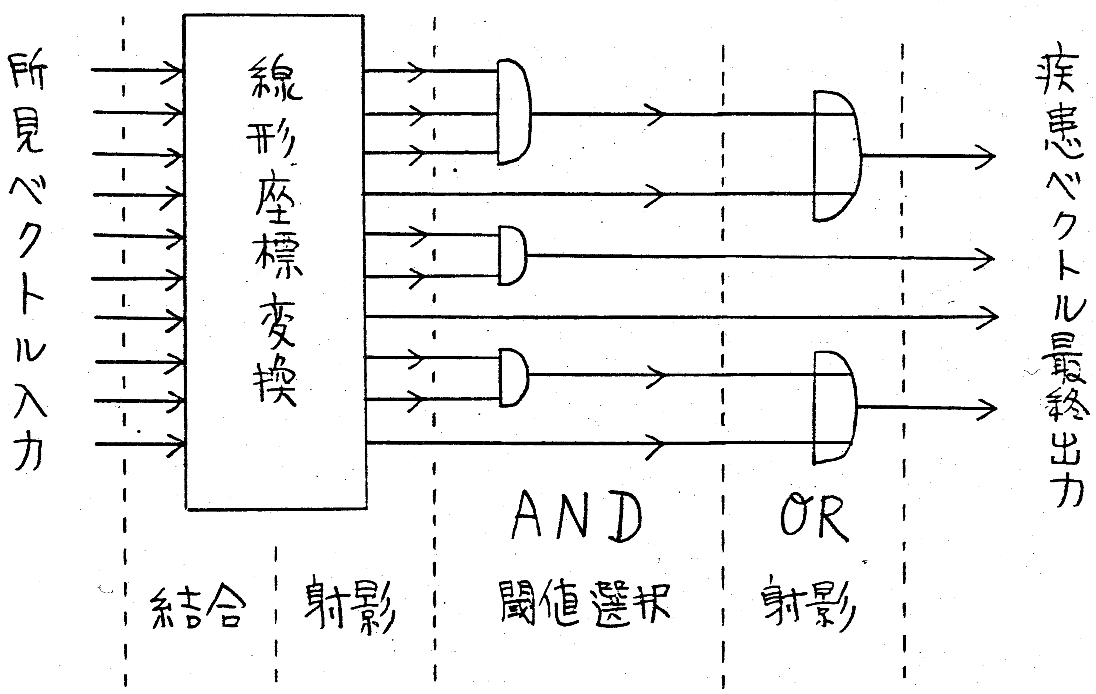

図2． 知識工学的応用例の構成概念

なおこれらの番号は連続自然数である必要はなく，両番号の対が同じ疾患の範囲で相互に識別可能であれば任意の文字列でよい。　また各変数の否定は対応タプルの重み係数(評点)にマイナス符号をつけて記述する。　上記の例では特定疾患に対して9タプル用意され，論理演算の結果1又は0タプルに集約される。　この機能は以下の手順で実現される。

① $\prod_{ik} G(c_{ijk}, i, j, k) \Rightarrow D(d_{ik}, i, k)$

② $\prod'_{ik} G(1, i, j, k) \Rightarrow W(w_{ik}, i, k)$

③ $D(d_{ik}, i, k) \underset{ik}{(\mathrm{H})} hW(w_{ik}, i, k) \Rightarrow E(e_{ik}, i, k)$

④ $\prod_{i} E(e_{ik}, i, k) \Rightarrow F(f_i, i)$

ここで①は射影により項内番号を消去し積項番号毎に重み係数の累計を与える。　②は関係表Gを初期設定した後，絶対値射影によって項内番号を消去して積項番号毎に可能最大しきい値を算出する。　③は閾値関係表を用いた閾値選択によりAND演算するが，しきい係数 h を1から減らしていくことにより，ある程度の条件不満足には目こぼしがきくFuzzy Andが実現出来る。　これは，あいまい性を必ず内在している現実世界や医療の判断を知識としてとりこむ際には，厳密な論理積関数よりもむしろ有用と思われる。　④は射影により積項番号を消去してOR演算し，疾患別に重み係数の累計を与えて目的とする疾患ベクトル表を出力する。

以上でゲートアレイとの対応づけを述べたが，重みづけ，閾値，タプル数が任意可変であるため非常に柔軟に機能する点に違いがある。

7、 知識工学的な適用例

今迄述べてきた拡張関係演算を用いて，バックトラックなしの確率的推論を行う場合の基本的構成例を述べる。
その処理の流れは，先ず入力の所見ベクトルBを相関知識表Aを用いて疾患ベクトルCへ線形変換する。 知識表AにはANDをとる範囲を規定する積項番号欄がありCにも出力される。 次にCから射影により所見属性欄（前述の項内番号の役割を果している）を消去しその重み係数を集約して関係表Dに出力する。
一方知識表Aを絶対値射影して所見属性欄（項内番号相当）を消去することにより可能最大閾値を関係表Wに算出する。
次にWを閾値表とし，hをそれにかかる閾係数としてDをAND演算しEを得る。 最後にDから射影により積項番号を消去してOR演算し結果Fを得る。

所見ベクトル表Bの投入作成は，それを容易にするために知識表Aを所見属性に絶対値射影してPを得て，その実現値をディスプレイしながら重み係数欄だけ新規投入させる

方法をとり，これを演算要素記号㊀で表わす。

以上の手順を演算要素記号を用いて以下に示す。

① $\Pi'_{i} A(a_{ijk}, i, j, k) \Rightarrow P(p_i, i)$

② $\Pi'_{jk} A(a_{ijk}, i, j, k) \Rightarrow W(w_{jk}, j, k)$

③ $\underset{i}{\textcircled{I}} P(p_i, i) \Rightarrow B(b_i, i)$

④ $B(b_i, i) \underset{i}{\bowtie} A(a_{ijk}, i, j, k) \Rightarrow C(c_{ijk}, i, j, k)$

⑤ $\Pi_{jk} C(c_{ijk}, i, j, k) \Rightarrow D(d_{jk}, j, k)$

⑥ $D(d_{jk}, j, k) \underset{jk}{\textcircled{H}} h\, W(w_{jk}, j, k) \Rightarrow E(e_{jk}, j, k)$

⑦ $\Pi_{k} E(e_{jk}, j, k) \Rightarrow F(f_k, k)$

ここでの属性欄の意味は例えば以下の通りである。

i ： 所見，ただし②⑤に於いては項内番号相当

j ： 積項番号

k ： 疾患

本実行に際しては知識表Aに更新がない限り③～⑦の5ステップを実行すればよい。

8． インプリメントとの対応

上述の演算要素はプログラムとしてRISS－Ⅱでインプリメントしてあるが，集合演算マシンとしてハードウエアインプリメントする仕様においても必要機能と考えられる。ただし重み係数の線形計算や閾値論理に関しては多少のバリ

エーションがあり得ること，そして結合演算に於ける内積，外積，直積を区別しまた内積を用いた割算，重み係数の正規化調整等を個別にインプリメントした方が効率的又は便利であることも考えられる。　またさらに，それらを組合せたマクロ演算の定義と走行が便利で効率的であるよう工夫する必要がある。　さらに今後の課題として，逆マトリックス生成に対応する関係表演算や事実データから知識を抽出生成するための多変量統計解析の実現に向けて，集合演算としての整備拡充を考える必要があろう。

9、あとがき

以上タプル重み係数としきい値を導入した関係演算とその応用について述べた。　これの意図する所は，ロジックプログラミングに代替することではなく，その配下で動作する大量知識ベースの，集合演算アーキテクチヤによる効率的な処理の実現にある。　また，厳密な論理式にはなじまない現実世界の，多要因・あいまい性・確率を伴った判断・演繹には，閾値論理で対応することで，重みづけとしきい値の変更やタプルの追加による論理関数の柔軟な変更を実現できるであろう。　生物の情報処理もこのように行われていると考えられる。

最後に，本検討を支援していただいた武蔵野電気通信研究所基礎部オ一研究室の山下室長始め室員の方々（並びにシステム20/20の円滑な運用）に感謝いたします。

［関連報告］

［１］　情処２３全大３Ｇ－６
（関係更新演算の機能及び定義と実行例）
［２］　信学技報ＡＬ８０－８４
（簡単な例題に対して関係演算でPROLOGをまねたもの）
［３］　情処２４全大４Ｇ－９
（連想メモリを用いた拡張形関係割算器の構成方法）
［４］　情処２２全大７Ｂ－４
（ＲＩＳＳ－Ⅱにおける先行投入応答抑止会話方式）
［５］　信学５７総全１２５４
（ＲＩＳＳ－Ⅱにおけるソーティング起動削減対策）