数理解析研究所講究録
522巻 1984年 295-307

結合節付加による不等号質問の処理

京大工学部　吉川 正俊 (Masatoshi Yoshikawa)
京大工学部　上林 弥彦 (Yahiko Kambayashi)

1. まえがき

データベースには数値など全順序が定義されているデータを格納することが多く，したがって不等号を含む質問もよく現われる。自然不等号質問は，質問最適化を考える上で関係質問の重要な部分クラスであり，その一部のものは準結合のみを用いて処理できることが知られている[BERNG81]が，一般のもの（質問グラフに閉路を持つものを含む）に関しては効率良い処理方法が知られていない。

本稿では，質問グラフが閉路を持つような自然不等号質問を処理するために閉路内の各枝に適当に結合条件式を付加して質問の変換を行ない，変換後の質問グラフに沿って一般化準結合を実行する方法について述べる。本稿で述べる方法は，自然結合質問の巡回型質問を処理するために我々が導入した属性付加による方法を特別な場合として含む。

2. 基本的事項

関係をR，対応する関係スキームを$\underline{R}$で表わす。一般にデータベース$D=(R_1, \cdots, R_n)$に対する質問は，条件式$\mathcal{g}$と目的属性集合から成るが，本稿では後者については考慮しない。$(\sigma_{\mathcal{g}}(R_1 \times \cdots \times R_n))[\underline{R_i}]$を$\mathcal{g}$に関する$R_i$の部分解と呼ぶ。本稿では，すべての関係の部分解を求める処理手順について論じる。

条件式$\mathcal{g}$が$R_i.A_k \ \theta \ R_j.A_l$（ただし$A_k \in \underline{R_i}$, $A_l \in \underline{R_j}$）の形の結合節の論理積で表わされる質問を不等号結合質問と言う。ただし，ここでθは比較演算子“$=$”,“$<$”,“$\leq$”,“$>$”,“$\geq$”のうちのいずれかであるとし，それぞれに対してθ^{-1}として“$=$”,“$>$”,“$\geq$”,“$<$”,“$\leq$”が定義されているものとする。

結合グラフ($J_{\mathcal{g}}$)は，不等号結合質問の条件式を表現する有向グラフであり，各節点は関係とその属性の対を表わし，結合節$R_i.A_k > R_j.A_l$ $(R_i.A_k \geq R_j.A_l)$に対応して有向枝$<R_i.A_k, R_j.A_l>$が存在する。ただし，結合節$R_i.A_k = R_j.A_l$は$(R_i.A_k \leq R_j.A_l) \wedge (R_i.A_k \geq R_j.A_l)$として解釈する。

自然不等号質問（NI質問）とは，不等号結合質問のうちその条件式が次の性質を満足するもののことを言う。

“質問中で参照される相異なる任意の2つの関係スキーム$\underline{R_i}$, $\underline{R_j}$に対し，$A \in \underline{R_i} \cap \underline{R_j}$が成立するときまたそのとき

に限り，結合グラフ内で節点 $R_i.A$ と $R_j.A$ は同一弱連結成分内にある。"

本稿ではNI質問を考察の対象とする。

質問グラフ (G_q) はNI質問の条件式を表現する無向グラフである。各節点は関係スキームを表わし，枝 $e_{ij}=\langle R_i, R_j\rangle$ は q において R_i と R_j の間で定義されている結合節の集合 $C_{ij}=\{c_{ij}^1, c_{ij}^2, \cdots, c_{ij}^k\}$ を表わす。ただし e_{ij} には C_{ij} に対応してラベル集合 $\{l_{ij}^1, \cdots, l_{ij}^k\}$ が付加されている。c_{ij}^t が $R_i.A_t\ \theta_t\ R_j.A_t$ とするなら $l_{ij}^t=\langle ja(l_{ij}^t), op(l_{ij}^t)\rangle \triangleq \langle A_t, \theta_t\rangle$ とし，$l_{ji}^t \triangleq (l_{ij}^t)^{-1} \triangleq \langle A_t, \theta_t^{-1}\rangle$，$L_{ji} \triangleq (L_{ij})^{-1} \triangleq \{(l_{ij}^1)^{-1}, \cdots, (l_{ij}^k)^{-1}\}$ と定義する。（ただし $1 \leq t \leq k$）。

3. 一般化準結合

本節では，[KAMBY8206] で導入した一般化準結合の概念を，不等号を含む質問も取り扱えるように拡張する。そのために，不等号射影，複数結合属性上の準結合の定義を行なう。

不等号を含む質問では，一般に処理途中の関係内に属性名は同じでもデータ値の異なる複数個の列が存在し得る。以後それらを陽に区別する必要がある場合は，各属性が最初にどの関係に存在していたかを示す上添字を付け，さらに区別が必要な場合には，上添字にその情報を付加する。

$A_1, \cdots, A_k$ を属性（上添字が付いていてもよい）の集まり，$\theta_1, \cdots, \theta_k$ を対応する比較演算子の集まりとするとき，$(A_1\theta_1, \cdots, A_k\theta_k)$ を射影条件式と言う。$A_1, \cdots, A_k$ を含む属性集合上の関係の２つの組 t, t' に対し，$\bigwedge_{i=1}^{k} (t[A_i]\ \theta_i\ t'[A_i])$ が成立するとき，t は t' より $(A_1\theta_1, \cdots, A_k\theta_k)$－小さいと言い，$t'$ は t より $(A_1\theta_1, \cdots, A_k\theta_k)$－大きいと言う。これにより，ある組集合が与えられたとき，組の間に半順序関係が定義でき，$(A_1\theta_1, \cdots, A_k\theta_k)$－極大（小）な組も定義することができる。この概念を用いて関係代数における射影を拡張した不等号射影を以下のように定義する。

$$R[A_1\theta_1, \cdots, A_k\theta_k] \triangleq \{t[A_1, \cdots, A_k] \mid t \text{ は } (A_1\theta_1, \cdots, A_k\theta_k)\text{-極小}\}$$

結合節 c_h : $R.A_h^{k_h}\ \theta_h\ S.B_h^{l_h}$ に対し，

$$att(c_h, R) \triangleq A_h^{k_h}, \quad att(c_h, S) \triangleq B_h^{l_h}$$

また，$C \equiv c_1 \wedge \cdots \wedge c_m$ に対し

$$att(C, R) \triangleq \bigcup_{h=1}^{m} att(c_h, R)$$

$$att(C, S) \triangleq \bigcup_{h=1}^{m} att(c_h, S)$$

とし，さらに射影条件式を結合条件式を用いて表現し，

$$R[C] \triangleq R[A_1^{k_1}\theta_1, \cdots, A_m^{k_m}\theta_m]$$

$$S[C] \triangleq S[B_1^{l_1}\theta_1^{-1}, \cdots, B_m^{l_m}\theta_m^{-1}]$$

と定義する。また，２つの組 $r \in R, s \in S$ が C を満足すると

き，それを $r\langle C\rangle s$ で表わす。

C を結合条件式 $c_1\wedge\cdots\wedge c_m$ とする。$att(C,R)\subseteq\underline{R}$，$att(C,S)\subseteq\underline{S}$ のとき，R と S の C 上の結合 $R\bowtie_C S$ は，

$$R\bowtie_C S \triangleq \{rs \mid r\in R, s\in S \text{ s.t. } r\langle C\rangle s\}$$

で定義される。また R の S による C 上の準結合 $R\ltimes_C S$ は

$$R\ltimes_C S \triangleq (R\bowtie_C S)[\underline{R}]$$

で定義される。一般に $R\ltimes_C S = R\ltimes_C S[C]$ が成立する。

また，$att(C,S)\subseteq\underline{S}$ のとき，R の S による C 上の一般化準結合 $R\,\bar{\ltimes}_C\,S$ は，

$$R\,\bar{\ltimes}_C\,S \triangleq R\ltimes_{C'} S[C]$$

で定義される。ただし，C' は $\{c_j \mid att(c_j,R)\in\underline{R}\}$ なる結合節の論理積結合であるとする。この定義より，結合節 $c: R.A\,\theta\,S.B$ に対し $A\notin\underline{R}$ ならば，任意の属性ラベル $A'(\notin\underline{R})$ に関して C' を結合節 $R.A'\,\theta\,S.B$ とするとき $R\,\bar{\ltimes}_C\,S = R\,\bar{\ltimes}_{C'}\,S$ が成立する。

4. 単純閉路質問の処理方法

本節では，3 節で定義した一般化準結合を用いて，質問グラフが単一の閉路のみで表わされるような単純閉路質問を処理する方法について述べる。任意の2つの関係間には質問グラフにおいて2つの経路が存在するが，質問処理の際はそれら関係間の組がいずれの経路に沿っても結合可能となること

を確認する必要がある。本節では，そのための方法として，(i)枝の埋め込みによる質問グラフの変換，(ii)一つの関係の結合属性の値に基づく付き合せ，について述べる。

我々は，理解を容易にするために，ある変換を実行した後の質問グラフに基づいて記述する。そこで，まず質問グラフをより厳密に定義し直しその上で質問グラフの変換について説明する。

質問グラフの各枝は，結合節を表わしているが，3節の初めに述べた理由により結合節 $R_i.A\ \theta\ R_j.A$ は厳密には $R_i.A^i\ \theta R_j.A^j$ を表わしていることになる。したがって，対応するラベルを l_{ij} とすると $l_{ij}=\langle(A^i, A^j), \theta\rangle$ となり，$l_{ij}^{-1}=\langle(A^j, A^i), \theta^{-1}\rangle$ とする。ただし，上記の結合節のように関係の添字と対応する属性の上添字が等しいとき（より厳密には，$R_i.A^k\ \theta\ R_j.A^l$ において $i=k$ かつ $j=l$ のとき）には，2節で定義したように対応するラベル l_{ij} は簡潔に $\langle A, \theta\rangle$ と表わすものとする。したがって，与えられたNI質問を表現する場合には2節の定義で十分であり，ここで定義した記法を用いる必要のあるラベルは，一般には質問グラフの変換によって生じた新しいラベルである。

4.1 枝の埋め込みによる質問グラフの変換

n 個の関係 $R_1, \cdots, R_n$ から成る質問の質問グラフが単一の閉

路によって表わされ，枝$\langle R_i, R_{i+1}\rangle$に対応するラベル集合を$L_i$ ($i=1,\cdots,n$ ただし $R_{n+1}\equiv R_1$) とする。(図1(a)参照) L_iによって表現される結合条件式をC_i ($i=1,\cdots,n$) とするとき，この方法は，C_nによって結合可能なR_1とR_nの組が，経路$R_1-R_2-\cdots-R_{n-1}-R_n$に沿っても結合可能か否か，また他の関係の組に関してはそれと結合可能なR_1とR_nの組どうしがC_nによって結合可能か否かを調べる。

自然結合質問における巡回型質問の処理と同様に，質問グラフのある1つの枝（ここでは$\langle R_n, R_1\rangle$とする）を他の枝に"埋め込む"．既ち，図1(b),(c)で表現されるように

$$L_i' \leftarrow L_i \cup L_n^{-1} \quad (i=1,\cdots,n-1)$$

とし，変換後の質問グラフに沿って順に一般化準結合

$$R_2 \underset{C_1'}{\ltimes} R_1,\quad R_3 \underset{C_2'}{\ltimes} R_2,\quad \cdots,\quad R_n \underset{C_{n-1}'}{\ltimes} R_{n-1} \qquad (i)$$

$$R_{n-1} \underset{C_{n-1}'}{\ltimes} R_n,\quad R_{n-2} \underset{C_{n-2}'}{\ltimes} R_{n-1},\quad \cdots,\quad R_1 \underset{C_1'}{\ltimes} R_2 \qquad (ii)$$

を実行することにより，すべての関係の部分解を得ることができる。（ただし，C_i'はL_i'によって表現される結合条件式とする。）実際の処理は上記の一般化準結合列(i)と(ii)をシャフルした任意の順序で行なってもよい。その場合，ある整数j ($1\leq j\leq n$) が存在し，一般に処理系列(i)の$R_{j+1} \underset{C_{j+1}'}{\ltimes} R_j$から最後までの一般化準結合($j=n$のときは空）と処理系列(ii)の

$R_{j-1} \underset{C_j}{\bowtie} R_j$ から最後までの一般化準結合（$j=1$のときは空）は，通常の準結合に置き換えても部分解を求める能力に関しては同等となる。R_jは最初に部分解を得ることのできる関係である。直観的には，上記の操作列(i)では各関係R_kの組が質問グラフ中の経路$R_k-R_{k-1}-\cdots-R_1$に沿って結合可能な$R_1[C_n]$の組を求めており，操作列(ii)では，各関係R_kの組が質問グラフ中の経路$R_k-R_{k+1}-\cdots-R_n$に沿って結合可能な$R_n[C_n]$の組を求めている。

図2(a)に示す質問グラフにおいて，枝$\langle R_1, R_5 \rangle$を埋め込む場合，先の(i),(ii)に対応する一般化準結合列は，それぞれ図2(b),(c)のように表現できる。$j=3$とすると，まず図2(d)に示される有向枝に沿って一般化準結合を実行し，次に図2(e)に示される有向枝に沿って一般化準結合を実行する。図2(d),(e)はそれぞれ巡回型質問の処理のための一般化準結合操作列UP・DOWN[KAMBY8206]に対応している。

4.2　一つの関係の結合属性の値に基づく付き合わせ

ここでは，質問グラフ内のある関係の一つの結合属性の値を，4.1の方法と同様に質問グラフの枝に沿って両方向に転送して行き，付き合わせを行なう方法について述べる。

まず例を用いて説明する。図2(a)の質問グラフにおいて，$R_5[A]$の値によって付き合わせを行なうものとする。例えば

R_3のある組t_3が部分解に含まれるか否かを調べるためには，経路$R_3-R_2-R_1-R_5$に沿ってt_3と結合可能な$R_5[A]$の値の集合と経路$R_3-R_4-R_5$に沿ってt_3と結合可能な$R_5[A]$の値の集合との共通集合が非空か否かを調べればよい。したがって，図３　に示すように質問グラフの各枝にラベル$L_{5=}:(A^5,=)$を付加することによって質問グラフを変換し，変換後の質問グラフの枝に沿って順に一般化準結合を行なえばよい。即ち，以下の処理系列(iii), (iv) の任意のシャフルを実行する。

$$R_1 \underset{A<,\,A^5=}{\boxtimes} R_5,\quad R_2 \underset{B=,\,A^5=}{\boxtimes} R_1,\quad R_3 \underset{C\geq,\,A^5=}{\boxtimes} R_2,\quad R_4 \underset{D>,\,A^5=}{\boxtimes} R_3,\quad R_5 \underset{E\leq,\,A^5=}{\boxtimes} R_4 \qquad \text{(iii)}$$

$$R_4 \underset{E\geq,\,A^5=}{\boxtimes} R_5,\quad R_3 \underset{D<,\,A^5=}{\boxtimes} R_4,\quad R_1 \underset{C\leq,\,A^5=}{\boxtimes} R_3,\quad R_1 \underset{B=,\,A^5=}{\boxtimes} R_2,\quad R_5 \underset{A>,\,A^5=}{\boxtimes} R_1 \qquad \text{(iv)}$$

ただし，(iv) の最後の操作は不要となる。

上記の例では，最初に与えられた質問の枝$\langle R_1, R_5\rangle$に付加されている演算子ラベルが“<”であるため，ある関係R_iの組t_iが経路$R_i-R_{i-1}-\cdots-R_1-R_5$に沿って$t_5[A]=a_0$なる$R_5$の組$t_5$と結合可能であれば，$t_i$は$t_5'[A]\geq a_0$を満足する任意の組$t_5'(\in R_5)$と結合可能となる。したがって，経路$R_i-R_{i-1}-\cdots-R_1-R_5$に沿って$t_i$と結合可能な$R_5[A]$の値の集合を知るためには，その集合の最小値（それをa_mとする）を知るだけで十分である。ゆえに，t_iが経路$R_i-R_{i+1}-\cdots-R_5$に沿って結合可能な$R_5[A]$の値集合のうち最大のものをa_Mとする

と，$a_m \leq a_M$ が成立するなら組 t_i は R_i の部分解に含まれることがわかる。(図4参照) したがって，(iii)，(iv) のかわりに以下の処理系列 (v)，(vi) を実行することによりすべての部分解を得ることができる。(ただし，この場合も (vi) の最後の操作は不要)

ただし，処理前に関係 R_5 には属性 A^{5M}，A^{5m} が付加されており，その値は $R_5[A]$ と同じとする。

一般的には，質問グラフが図1(a)のように与えられている場合，L_n に属する各ラベル l_n^a を表1のように変換して得られるラベル $l_n^{a'}$ の集合を L_n' とし，

$$L_i' \leftarrow L_i \cup L_n' \quad (i=1, \cdots, n)$$

とした後，変換後の質問グラフの経路 $R_n - R_1 - R_2 - \cdots - R_n$ 及び $R_n - R_{n-1} - \cdots - R_1$ に沿って順に一般化準結合を実行する。

謝辞　日頃御指導頂く矢島脩三教授をはじめとする研究室の皆様に深謝します。

参考文献

[BERNG81] Bernstein,P.A. and Goodman,N., "The Power of Inequality Semijoins", Information Systems, Vol.6, No.4, pp.255-265, 1981.

[KAMBY8206] Kambayashi,Y., Yoshikawa,M. and Yajima,S., "Query Processing for Distributed Databases Using Generalized Semi-Joins", Proc. of ACM SIGMOD International Conference on Management of Data, pp.151-160, June 1982.

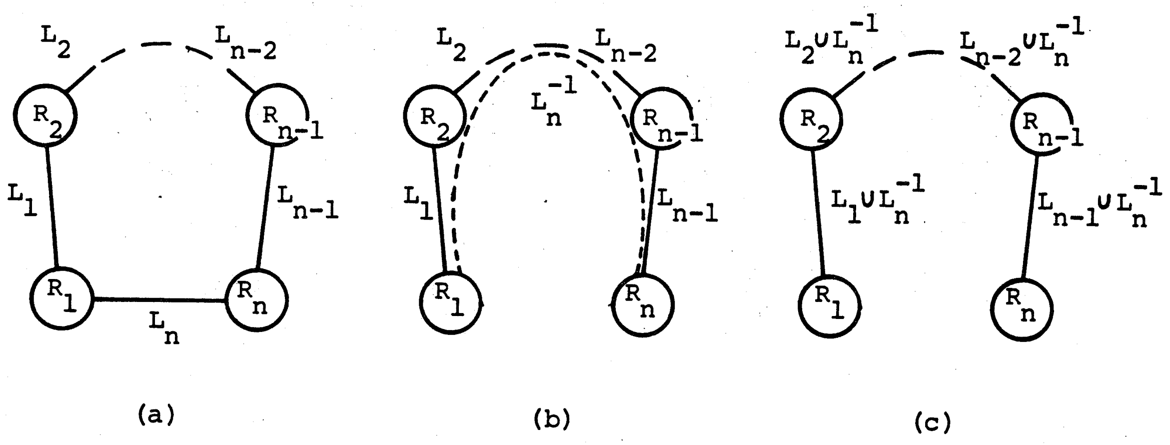

図1. 枝の埋め込みによる閉路の除去

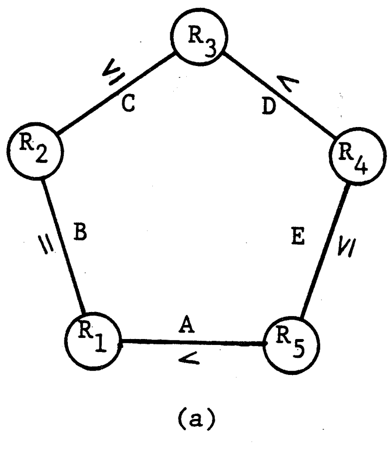

図2. 結合節付加の例

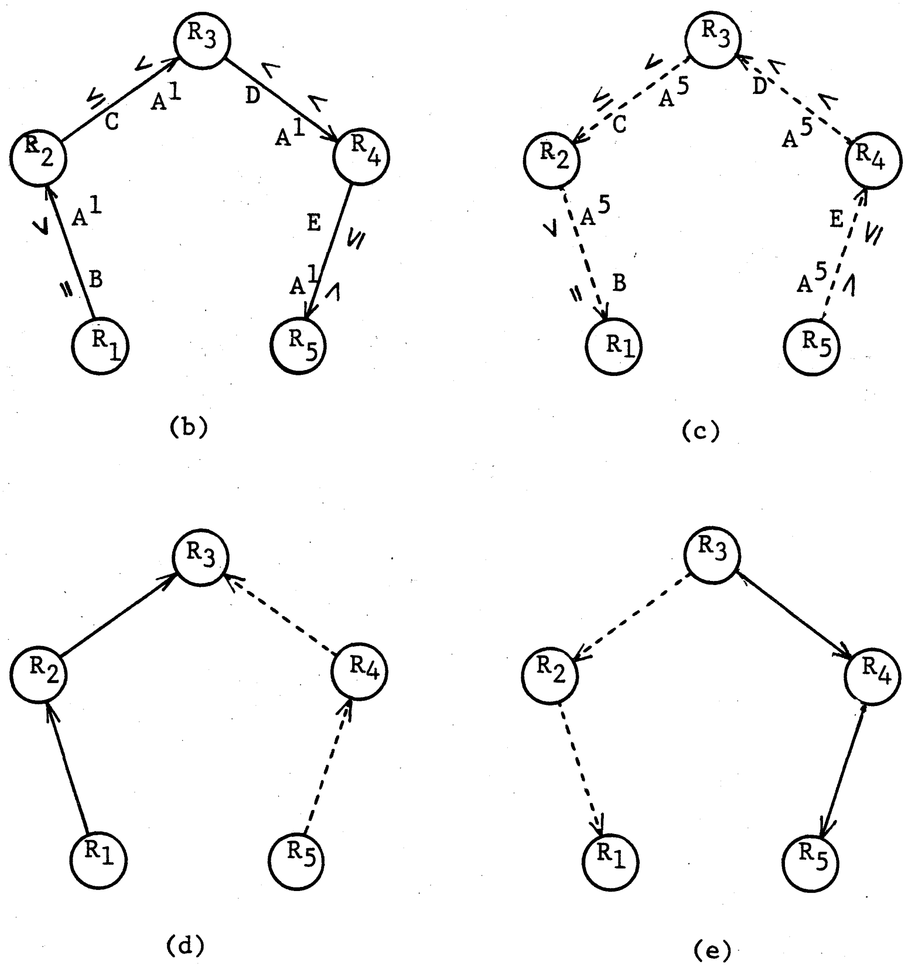

図2.（続き）

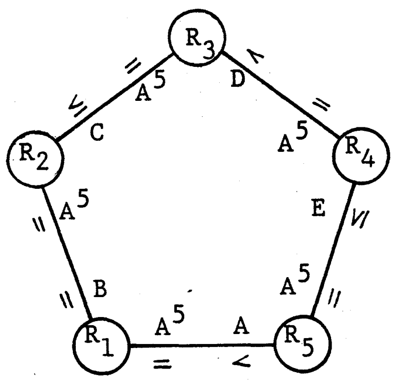

図3 属性 A^5 の等号による付き合わせ

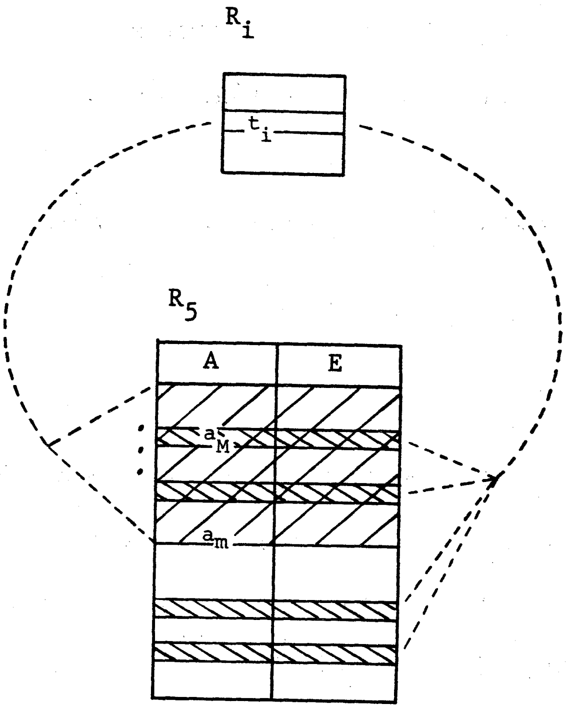

図4 組t_iと結合可能なR_5の組

l_n^h	$l_n^{h'}$
$A_i, \leq$	$(A_i^{nM},\ A_i^{nm}), \geq$
$A_i, \geq$	$(A_i^{nM},\ A_i^{nm}), \leq$
$A_i, =$	$(A_i^{nM},\ A_i^{nm}), =$
$A_i, <$	$(A_i^{nM},\ A_i^{nm}), \geq$
$A_i, >$	$(A_i^{nM},\ A_i^{nm}), \leq$

表1. 属性A_i^nの値による付き合わせを行なう場合に付加すべきラベル