数理解析研究所講究録 1325 巻 2003 年 110-115



# 正規パターン言語の集合積の包含問題について
# Containment Problem for Intersections of Regular Pattern Language

四天王寺国際仏教大学 寺田 幹治*(Mikiharu Terada)
大阪府立大学総合科学部数理・情報科学科 佐藤 優子**(Masako Sato)
*International Buddhist University
**Department of Mathematics and Information Sciences
College of Integrated Arts and Sciences, Osaka Prefecture University

keyword : 包含問題, 正規パターン言語

## 1 はじめに

パターンとは, 定数記号と変数 (記号) からなる空でない有限文字列である。各変数に空でない定数列を代入して得られる定数文字列からなる言語をパターン言語という。パターン言語のクラスは, 正例から帰納推論可能な言語族として, Angluin[1] によって導入された。

本稿では, 2 つの正規パターン $p, q$ で定義される言語の共通部分 $L = L(p) \cap L(q)$ の包含問題を扱う。この共通部分は, $p, q$ の極大例化とよばれるパターンの言語和として表すことができる。一般のパターンでは, 極大例化は, 必ずしも有限ではない。たとえば, $p = xax, q = axa$ とすると, $a^n$ はすべて極大例化となる。一方, パターンが正規ならば, Sato and Mukouchi[3] の結果から, 極大例化 $r$ の長さは, 高々 $p$ と $q$ の長さの和で抑えられ, 有限である。しかし, すべての極大例化を求める問題は, 計算可能ではあるが, かなりの計算量が予想される。なぜなら, 変数への空列代入を許す erasing 正規パターンに関しては, その問題は, NP 困難という結果が得られているからである。本稿では, $p, q$ が同じ長さ $n$ をもつパターンに関して, その言語の共通部分の包含問題が, 長さが $n+1$ の極大例化集合の関係を調べることによって, 解決できることを示す。このような極大例化集合を特定するために, パターンに対して, 分割点及び分割集合の概念を導入する。

## 2 正規パターン言語

$\Sigma$ を少なくとも 2 個の定数記号を含むアルファベットとし, $X$ を変数記号からなる加算集合とする。ただし, $\Sigma$ と $X$ は互いに排反とする。

パターンとは, $\Sigma \cup X$ 上の空でない文字列である。同じ変数が高々 1 回しか現れないパターンを正規パターンという。

パターン $p$ の長さを $|p|$, パターン $p$ の $i$ 番目の記号を $p[i]$ で表す。また, $i \leq j$ に対して, $p[i]$ から $p[j]$ までの長さ $j-i+1$ の $p$ の部分列を $p[i,j]$ で表す。さらに, $I_v(p) = \{i \mid p[i] \in X\}$, $I_c(p) = \{i \mid p[i] \in \Sigma\}$ と表すことにする。たとえば, $p = abxcyde$ ならば, $p[2] = b$, $p[5] = y$, $I_c(p) = \{1,2,4,6,7\}$, $I_v(p) = \{3,5\}$ である。

代入とは, すべての定数をそれ自身に写すパターンからパターンへの準同型写像をいう。代入 $\theta$ による $p$ の像を $p\theta$ で表す。以下, 代入は, 非消去代入, すなわち, $|x\theta| \geq 1$ $(x \in X)$ を満たす代入とする。

パターン $q$ がパターン $p$ の汎化, または $p$ が $q$ の例化であるとは, $q\theta = p$ を満たす代入 $\theta$ が存在することをいい, $p \preceq q$ で表す。特に, $p \preceq q$ かつ $q \not\preceq p$ のとき, $p \prec q$ で表す。明らかに, $p \preceq q$ な

らば，$|p| \geq |q|$ である。

パターン $p$ に対して，$\Sigma$ 上の言語 $L(p) = \{w \in \Sigma^+ \mid w \preceq p\}$ を定義する。$p \preceq q$ ならば，$L(p) \subseteq L(q)$ である。従って，$p \preceq q$ かつ $q \preceq p$ ならば，$L(p) = L(q)$ が成り立つ。このようなパターン $p$ と $q$ は**等価**とよばれ，変数名を除いて一致する。本稿では，等価なパターンは同一視する。$\Sigma$ 上の言語 $L$ は，$L = L(p)$ を満たすパターン $p$ が存在するとき，**パターン言語**とよばれる。

以下，本項では，正規パターンのみを扱い，正規を略し，単に，パターンとよぶことにする。正規パターンの全体を $\mathcal{RP}$ で表す。また，長さ $n$ の正規パターンの集合を $\mathcal{RP}_n$ で表す。

パターン $p$ がパターン集合 $P$ の**汎化**とは，$p$ が $P$ に含まれるすべてのパターンの汎化であることをいう。$p$ が $P$ の汎化で，$q \prec p$ となる $P$ の汎化 $q$ が存在しないとき，$p$ は $P$ の**極小汎化** (minimal generalization; mg と略す) という。また，$P$ の極小汎化の中で最長のパターンを $P$ の**最長極小汎化** (longest minimal generalization; lmg と略す) という。$P$ の mg, lmg となるパターンの集合をそれぞれ，mg $P$，lmg $P$ で表す。

パターン集合 $P$ に対して，その**例化**，**極大例化** (maximal instance; mi と略す) および**最短極小例化** (shortest maximal instance; smi と略す) を同様に定義する。mi $P$ 及び smi $P$ の記号も同様に定義する。

パターン $x$ は任意のパターンの汎化であるので，任意の $P$ に対して，$P$ の汎化は必ず存在するが，その例化は存在するとは限らない。たとえば $P = \{ax, bx\}$ $(a \neq b)$ の例化は存在しない。

$p \in$lmg $P$ ならば $|p| = \min\{|q| \mid q \in P\}$ である。また，$P$ が等長パターンの集合ならば，lmg $P$ は一意に定まる。パターン $p$ が等長パターン集合 $P$ の lmg ならば，lmg $P$ は $p$ のみの集合であるので，単に，lmg $P = p$ とかく。

パターンの例化および汎化については次のような結果が得られている。

**補題 2.1** **(Mukouchi[2])** $n = \min\{|p'| \mid p' \in P\}$ とするとき，$p$ が長さ $n$ の $P$ の汎化ならば，$q \preceq p$ となるパターン $q \in$lmg $P$ が存在する。

**補題 2.2** **(Mukouchi[2])** 集合 $\mathcal{RP}_n$ は，関係 $\preceq$ に関して有限束である。ただし，任意の $p \in \mathcal{RP}_n$ に対して，$\phi \preceq p$ とする。

上記の結果より，パターン $p$ が等長パターン集合 $P$ の smi ならば，lmg と同様に，smi $P = p$ とかく。

**補題 2.3** **(Mukouchi[2])** $|\Sigma| \geq 3$ とするとき，$L(p) \subseteq L(q)$ であることと $p \preceq q$ であることは同値である。

**補題 2.4** **(Shinohara[4])** 任意のパターン $p, q$ に対して，$p \preceq q$ であるか否かの決定問題は，$O(|p| + |q|)$ の時間で計算可能である。

本稿では，2 つのパターン言語の共通部分の包含問題を扱う。パターン $p, q$ に対して，

$$L(p, q) = L(p) \cap L(q)$$

とおく。次の結果は，mi の定義から直ちに示される。

**補題 2.5** パターン $p, q$ に対して，

$$L(p, q) = \bigcup_{r \in \mathrm{mi}\{p,q\}} L(r)$$

が成立する。

補題から次の結果も成り立つ。

**定理 2.6** $p_i, q_i$ $(i = 1, 2)$ をパターンとする。このとき，$L(p_1, p_2) \subseteq L(q_1, q_2)$ であることと

$$\forall r \in \mathrm{mi}\ \{p_1, p_2\}, \exists s \in \mathrm{mi}\ \{q_1, q_2\}\ \text{s.t.}\ r \preceq s$$

であることは同値である。

**補題 2.7** **(Sato and Mukouchi[3])** $r \in$mi $\{p, q\}$ とする。このとき，$|r| \leq |p| + |q|$ が成り立つ。

補題 2.4 及び 2.7 から，mi $\{p, q\}$ を求める問題は計算可能である。また，定理 2.6 より，パターン言語の共通部分の包含問題も，計算可能であることがわかる。

## 3 等長パターンの極大例化

パターン $p$ に対して，変数から定数記号または定数記号から変数に変わった後の位置を $p$ の**分割点**という。たとえば，$p = abx_1x_2bcx_3$ とすると，3, 5, 7 は $p$ の分割点である。$p$ の分割点の集合を $d_p$ で表す。$d_p = \{l_1, \cdots, l_k\}$ とすると，明らかに，$1 < l_i \le |p|, l_i \ne l_j$ $(1 \le i, j \le k,\ i \ne j)$ である。各 $i$ に対して，$p[l_i - 1] \in \Sigma$, $p[l_i] \in X$ であるか，$p[l_i-1] \in X, p[l_i] \in \Sigma$ である。前者を **$cv$分割点**，後者を **$vc$分割点**という。また，分割点ではない位置の文字が定数のとき，**$c$連続点**といい，変数のとき，**$v$連続点**という。$l$ が $c$ 連続点ならば，$p[l-1], p[l] \in \Sigma$ であり，$v$ 連続点ならば，$p[l-1], p[l] \in X$ である。以下，変数列を $\chi, \chi_1, \cdots$ 等で表す。また，パターン $p$ と $1 < l_1 < l_2 < \cdots < l_k \le |p|$ を満たす集合 $d = \{l_1, l_2 \cdots, l_k\}$ に対して，$p$ の部分列 $p_1 = p[1, l_1 - 1]$, $p_{i+1} = p[l_i, l_{i+1} - 1]$ $(i = 1, \cdots, k-1)$, $p_{k+1} = p[l_k, |p|]$ の連接で表示した式 $p = p_1p_2 \cdots p_kp_{k+1}$ を **$d$-表現**といい，$p_i$ を ($d$ における)$i$ 番目の部分列という。

$p, r$ を $r \preceq p$ を満たす等長パターンとする。このとき，

(1) $l \in d_r - d_p$ ならば，$l$ は $p$ の $v$連続点である。

(2) $l \in d_p - d_r$ ならば，$l$ は $r$ の $c$連続点である。

$w \in \Sigma^+$ を $w \preceq p$ を満たす $p$ と等長の定数列とする。このとき，smi $\{p, q\} = w$ を満たす等長なパターン $q$ は一般には一意ではない。しかし，$d_p = d_q$ も同時に満たすパターン $q$ は一意である。たとえば，$w = aabbbcc, p = xabxxxc$ に対して，smi $\{p, q\} = w$ を満たすパターン $q$ は，$aA_2A_3bbcA_7$ $(A_2 = a \text{ or } x,\ A_3 = b \text{ or } x,\ A_7 = c \text{ or } x)$ の 8 通りがある。しかし，これらのうち，$d_p = \{2, 4, 7\}$ をもつパターンは，$q = axxbbcx$ のみである。このようなパターン $q$ を $p$ の $w$ に関する**相補パターン**とよび，$p_w$ で表す。この定義から，明らかに，smi $\{p, p_w\} = w$, lmg $\{p, p_w\} = \chi$ となる。

パターン $p$ に含まれる変数 $x$ に長さ 2 の変数 $xx'$ を代入して得られるパターン $p\{x := xx'\}$ を $p$ の **1-expansion** とよぶ。

$p = w_1\chi_1w_2\chi_2 \cdots w_k\chi_kw_{k+1}$ $(w_1, \cdots, w_{k+1} \in \Sigma^+)$ のとき，$k$ 個の異なる 1-expansion $w_1\chi_1w_2 \cdots w_i(\chi_ix)w_{i+1} \cdots w_k\chi_kw_{k+1}$ $(i = 1, \cdots, k)$ が存在する。

長さ $n$ の等長パターン $p, q$ に対して，長さ $n+1$ の極大例化の集合を mi $\{p, q\}_{n+1}$ で表す。

$\xi \in$mi $\{p, q\}_{n+1}$ ならば，$\xi \preceq p' \preceq p, \xi \preceq q' \preceq q$ を満たす $p, q$ の 1-expansion $p', q'$ がそれぞれ存在し，$\xi =$smi $\{p', q'\}$ が成り立つ。

文字列 $w$ に対して，head$(w)$, tail$(w)$ で $w$ の先頭の文字，最後尾の文字を表すものとし，$\underline{w}$ を $w$ の先頭の文字を除いた文字列とし，$\overline{w}$ を $w$ の最後尾の文字を除いた文字列とする。従って，$w = \text{head}(w)\underline{w} = \overline{w}\text{tail}(w)$ が成り立つ。文字列の対 $(w, w')$ に対して，$\text{tail}(w) = \text{head}(w')$ ならば，$w\underline{w'} = \overline{w}w'$ が成り立つ。この文字列を $[w, w']$ で表す。明らかに，$|[w, w']| = |w| + |w'| - 1$ である。また，$w$ の $d = \{l_1, \cdots, l_k\}$**-分割表現** $w = w_1w_2 \cdots w_kw_{k+1}$ に対して，次のパターン集合を定義する：

$$
\begin{aligned}
Q_{w,d} = \{ & \\
& w_1w_2 \cdots w_i(x[w_{i+1}, w_{i+2}])(x[w_{i+3}, w_{i+4}]) \cdots \\
& (x[w_{j-2}, w_{j-1}])xw_jw_{j+1} \cdots w_{k+1} \mid \\
& \quad 1 \le i < j \le k,\ j - i + 1 \text{ は偶数}, \\
& \quad \text{tail}(w_s) = \text{head}(w_{s+1}),\ i + 1 \le s \le j - 2, \\
& \quad \text{tail}(w_{i-1}) \ne \text{head}(w_{i-1}) \\
& \qquad \text{または } \text{tail}(w_{i-1}) \ne \text{head}(w_i), \\
& \quad \text{head}(w_{j+1}) \ne \text{tail}(w_{j+1}) \\
& \qquad \text{または } \text{head}(w_{j+1}) \ne \text{tail}(w_j)\}.
\end{aligned}
$$

$n = |w_1 \cdots w_{k+1}|$ とすると，上記の集合に含まれるパターンの長さは，すべて $n+1$ である。$Q_{w,d}$ は，位置 $l_i, l_{i+2}, \cdots, l_{j-3}, l_{j-1}$ に $(j - i + 1)/2$ 個の変数をもち，上記の条件を満たすパターンの集合である。たとえば，$w = a^{10}$, $d = \{3, 4, 6, 9\}$ のとき，$Q_{w,d} = \{a^2(xa^2)xa^5, a^3(xa^4)xa^2\}$ である。

**補題 3.1** $w \in \Sigma^n, p \in \mathcal{RP}_n$ とする。$w \preceq p$ ならば，mi $\{p, p_w\}_{n+1} = Q_{w,d_p}$ が成立する。

(証明) $d_p = \{l_1, \cdots, l_k\}$ とする。$k$ が奇数の場合について証明する。偶数の場合も同様である。$w$ の $d_p$-表現を $w = w_1 \cdots w_{k+1}$ とおく。一般性を失うことなく，$p, p_w$ の $d_p$-表現は $p = w_1\chi_2 \cdots w_k\chi_{k+1}$, $p_w = \chi_1w_2 \cdots \chi_kw_{k+1}$ とおける。

(i) $\xi \in$mi $\{p, p_w\}_{n+1}$ とすると，$\xi =$smi $\{p', p'_w\}$ を満たす $p, p_w$ の 1-expansion $p', p'_w$ が存在する。$p' = w_1\chi_2 \cdots w_{i-1}(\chi_ix)w_{i+1} \cdots \chi_{k+1}$, $p'_w =$

$\chi_1 w_2 \cdots w_{j-1}(x\chi_j)w_{j+1}\cdots w_{k+1}$ とする。ただし，$i$ は偶数で，$j$ は奇数である。今，$i<j$ と仮定すると，$\xi=$smi $\{p', p'_w\}$ であるので，

(1) $j=i+1$ の場合は，$\xi=w_1\cdots w_i x w_{i+1}\cdots w_{k+1}$

(2) $j>i+1$ の場合は，$\xi=w_1\cdots w_i(x[w_{i+1},w_{i+2}])(x[w_{i+3},w_{i+4}])\cdots(x[w_{j-2},w_{j-1}])xw_j\cdots w_{k+1}$

でなければならない。ただし，(2) では，$\mathrm{tail}(w_s)=\mathrm{head}(w_{s+1})$ $(s=i+1,\cdots,j-2)$ である。このパターンは，$(j-i+1)/2$ 個の変数を持ち，それらの位置は $(l_i,l_{i+2},\cdots,l_{j-3},l_{j-1})$ である。もし，$\mathrm{tail}(w_{i-1})=\mathrm{head}(w_i)$ とすると，$p$ の変数列 $\chi_{i-2}$ を 1-expansion したパターンと $p'_w$ の最短極大例化は，$\eta=w_1\cdots w_{i-2}(x[w_{i-1},w_i])(x[w_{i+1},w_{i+2}])\cdots(x[w_{j-2},w_{2j-2}])xw_j\cdots w_{k+1}$ となり，$w_{i-1}w_i\preceq x[w_{i-1},w_i]$ ならば，$\xi\prec\eta$ を満たし，$\xi$ が極大例化であることに反する。この関係が満たされるのは，$\mathrm{head}(w_{i-1})\underline{w_{i-1}}w_i\preceq x\overline{w_{i-1}}w_i$ より，$\underline{w_{i-1}}=\overline{w_{i-1}}$ の場合であり，Uemura[6] より，$w_{i-1}$ は，同一の文字からなる定数列に限られる。従って，$\mathrm{head}(w_{i-1})=\mathrm{tail}(w_{i-1})$ となり矛盾する。

$j\le i$ の場合も同様である。したがって，長さが $n+1$ の $p$ と $p_w$ の極大例化は，$Q_{w,d_p}$ に含まれる。

(ii) 次に，$Q_{w,d_p}$ のパターンが $p, p_w$ の例化であることは明らかなので，極大であることを示す。$\xi\in Q_{w,d_p}$ が次式で表される場合のみを示す。パターン $\xi=w_1\cdots w_{2i}(x[w_{2i+1},w_{2i+2}])(x[w_{2i+3},w_{2i+4}])\cdots(x[w_{2j-3},w_{2j-2}])xw_{2j-1}\cdots w_{k+1}$ が極大例化ではないとすると，$\xi\prec\eta$ を満たす $\eta\in$mi $\{p,p_w\}$ が存在する。$\xi\prec\eta$ より，$n+1=|p|\ge|\eta|$ である。$|\eta|=n$ ならば，$\eta=$smi $\{p,p_w\}$ となるので，$\eta=w$ となり，矛盾である。故に，$|\eta|=n+1$ である。(i) の証明から，$\eta\in Q_{w,d_p}$ である。$\xi$ は，位置 $l_{2i},l_{2i+2},\cdots,l_{2j-2}$ で変数をもつので，$\eta$ もこれらの位置で変数をもつ。$\xi\prec\eta$ であるので，$l_{2i-2}$ または，$l_{2j}$ は変数である。$w_{2i-1}$, $w_{2j}$ の head, tail に関する条件から，このような $\eta$ は，$Q_{w,d_p}$ には含まれない。 ■

上記の結果より，$p$ の各分割点 $l\in d_p$ に対して，位置 $l$ に変数をもつ極大例化 $\xi\in$mi $\{p,p_w\}_{n+1}$ が唯ひとつ存在することがわかる。さらに，極大例化の変数の位置は，そのような分割点以外では起こらない。

等長パターン $p,q\in\mathcal{RP}_n$ に対して，$d_p\cup d_q$ を $(p,q)$ の分割集合といい，$d_{p,q}$ で表す。たとえば，

$$p=x_1\ a_2\ x_3x_4\ a_5\ a_6\ x_7x_8,$$
$$q=a_1\ a_2\ x_3x_4\ x_5\ a_6\ x_7x_8$$

とすると，$d_p=\{2,3,5,7\}$, $d_q=\{3,6,7\}$, $d_{p,q}=\{2,3,5,6,7\}$ である。

$r\preceq s$ を満たすパターン $r,s\in\mathcal{RP}_n$ に対して，集合 $P_{r,s}$ 及び $P_r$ を次のように定義する:

$$P_{r,s}=\{(p,q)\mid p,q\in\mathcal{RP}_n,\ r=\mathrm{smi}\ \{p,q\},\ s=\mathrm{lmg}\ \{p,q\}\},$$
$$P_r=\bigcup_{s\in\mathcal{RP}_n \mathrm{s.t.} r\preceq s} P_{r,s}.$$

$(p,q)\in P_{r,s}$ に対して，$d_{r,s}=\{l_1,l_2,\cdots,l_k\}$ とし，$r,s,p,q$ の $d_{r,s}$-表現における $i$ 番目の部分列をそれぞれ，$r_i,s_i,p_i,q_i$ とすると，

(1) $(p_i,q_i)\in P_{r_i,s_i}$.

(2) $r_i$ が変数列ならば，$s_i,p_i,q_i$ はすべて変数列である。

(3) $s_i$ が定数列ならば，$r_i,p_i,q_i$ はすべて定数列である。

(4) $r_i$ が定数列 $(=w_i)$ であり，$s_i$ が変数列ならば，$p_i,q_i$ は互いに $w_i$ に関して相補的である。

上記より，$p,q$ の値が異なる位置は，(4) の $r_i$ が定数列で $s_i$ は変数列となる位置のみである。このような部分列が出現する位置の集合を $\hat{d}_{r,s}$ で表す。すなわち，

$$\hat{d}_{r,s}=\{l\in d_{r,s}\mid r[l]\in\Sigma,\ s[l]\in X\}.$$

また，上記より，$r\preceq p\preceq s$ を満たすパターン $p$ に対して，$(p,q)\in P_{r,s}$ となるパターン $q$ がただ一つ存在する。従って，$(p,q)$ は順序対であるが，以下，$(p,q)$ と $(q,p)$ を同一視する。たとえば，$r=xabcc$, $s=xxbxx$ とすると，$P_{r,s}=\{(r,s),(xxbcx,xabxx),(xabcx,xxbxc),(xabxc,xxbcx)\}$ である。

**補題 3.2** $(p,q)\in P_{r,s}$ とする。このとき，

(1) $d_{p,q}=d_{r,s}\cup(d_p\cap d_q)$,

(2) $r=w\in\Sigma^+$ ならば，$d_s\cap(d_p\cap d_q)=\phi$ である。

(3) $s=\chi\in X^+$ ならば，$d_r\cap(d_p\cap d_q)=\phi$ である。

(証明) (1) $l \in d_{p,q}$ とする。$l \in d_p, l \notin d_q$ ならば，$l$ は，$q$ の $c$ 連続点であるか，または，$v$ 連続点である。前者の場合，$l$ は $s$ の切断点となり，後者の場合は $r$ の切断点である。従って，$l \in d_{r,s}$ が成り立つ。

次に，$l \in d_{r,s}$ とする。$l \in d_r$ とし，$r$ の $cv(vc)$ 切断点とすると，$r =$smi $\{p,q\}$ であるので，$l$ は，$p$ または $q$ の $cv(vc)$ 切断点でなければならない。$l \in d_s$ の場合も同様である。よって，$d_{r,s} \subseteq d_{p,q}$ である。(2), (3) は明らかである。 ■

上記の (1) より，$(p,q) \in P_r$ ならば，$d_r \subseteq d_{p,q}$ が成り立つ。

**補題 3.3** $(p,q) \in P_r$ とする。このとき，
(1) $\xi$ を $p,q$ の長さ $n+1$ の例化とする。$\xi[l] \in X$ ならば，$l$ は，$p$ 及び $q$ の $c$ 連続点ではない。
(2) $l \in d_p \cap d_q$ ならば，$\xi[l] \in X$ を満たす $\xi \in$mi $\{p,q\}_{n+1}$ が存在する。

(証明) (1) $l$ が $p$ の $c$ 連続点とする。$\xi \preceq p$ より，$\xi \preceq p'$ となる $p$ の 1-expansion $p'$ が存在する。明らかに，$p'[l] = p[l-1]$ または，$p'[l] = p[l]$ のいずれかである。いずれにしても，$p'[l]$ は変数にはならないので，$\xi \preceq p'$ に矛盾である。

(2) $l \in d_p \cap d_q$ とする。$l$ が $p,q$ の $cv$ 連続点 ($vc$ 連続点) ならば，変数 $p[l], q[l] (p[l-1], q[l-1])$ の 1-expansion の極大例化は，位置 $l$ で変数となる。$l$ が，$p$ の $cv$ 連続点で $q$ の $vc$ 連続点のとき，すなわち，$p[l-1,l] = ax$, $q[l-1,l] = yb$ $(a,b \in \Sigma)$ とする。$p,q$ の変数 $x,y$ の 1-expansion を $p',q'$ とすると，$r[1,l-2]axbr[l+1,n]$ は，$p',q'$ の長さ $n+1$ の例化である。従って，$r[1,l-2]axbr[l+1,n] \preceq \xi$ を満たす $\xi \in$mi $\{p,q\}_{n+1}$ が存在する。明らかに，$\xi[l] \in X$ である。 ■

## 4 包含問題と切断集合

この節では，等長パターン言語の共通部分の包含関係と，前節で導入した分割集合との関係について論じる。

$(p,q) \in P_r$ とする。このとき，$L(r) \subseteq L(p) \cap L(q)$ が成り立つ。更に，これらの言語の最短文字列集合は等しい。すなわち，

$$L(r) \cap \Sigma^n = L(p) \cap L(q) \cap \Sigma^n.$$

**定理 4.1** $(p_1,p_2),(q_1,q_2) \in P_r$ とする。
(1) $L(p_1,p_2) \subseteq L(q_1,q_2)$ ならば，$d_{p_1} \cap d_{p_2} \subseteq d_{q_1,q_2}$ である。
(2) $(p_1,p_2),(q_1,q_2) \in P_{r,s}, L(p_1,p_2) \subseteq L(q_1,q_2)$ ならば，$d_{p_1,p_2} \subseteq d_{q_1,q_2}, d_{p_1} \cap d_{p_2} \subseteq d_{q_1} \cap d_{q_2}$ である。

(証明) (1) $d_{p_1} \cap d_{p_2} \not\subseteq d_{q_1,q_2}$ とする。このとき，$l \in d_{p_1} \cap d_{p_2} - d_{q_1,q_2}$ が存在する。明らかに，$l$ は，$q_1(q_2)$ の $c$ 連続点であり，かつ $q_2(q_1)$ の $v$ 連続点であることを意味する。補題 3.3(2) より，$\xi[l] \in X$ を満たす $\xi \in$mi $\{p_1,p_2\}_{n+1}$ が存在する。仮定より，$L(p_1,p_2) \subseteq L(q_1,q_2)$ であるので，$\xi \preceq \eta$ を満たす $\eta \in$mi $\{q_1,q_2\}_{n+1}$ が存在しなければならない。明らかに，$\eta[l]$ は変数である。補題 3.3(1) より，$l$ は，$q_1,q_2$ の $c$ 連続点ではない。$l$ のとり方から，$q_1$ と $q_2$ のいづれかでは，$c$ 連続点であるので矛盾である。

(2) パターン対が $P_{r,s}$ に含まれるならば，補題 3.2(1) より，

$$d_{p_1,p_2} = d_{r,s} \cup (d_{p_1} \cap d_{p_2}),$$
$$d_{q_1,q_2} = d_{r,s} \cup (d_{q_1} \cap d_{q_2})$$

である。(1) の結果を用いると，$d_{p_1,p_2} \subseteq d_{q_1,q_2}$ が得られる。

次に，$l \in d_{p_1} \cap d_{p_2} - d_{q_1} \cap d_{q_2}$ とする。$l \in d_{q_1,q_2}$ であるので，$l \in d_{q_1} - d_{q_2}$ とする。すると，(1) と同様な方法で矛盾を示すことができる。 ■

定数列 $w = a_1a_2 \cdots a_n$ に対して，$p$ を定数と変数が交互に出現し，$w \preceq p$ となるパターンとする。このようなパターンは丁度 2 つ存在し，互に $w$ に関して相補的パターンである。これらのパターン対を $w$ の**最分割汎化対**という。$w = abcd$ ならば，対 $(axcy, xbyd)$ は $w$ の最分割汎化対である。$(p,q)$ が $w$ の最分割汎化対であるならば，$(p,q) \in P_{w,X}$ であり，最初の位置を除いて，すべての位置が，$d_p (= d_q)$ に含まれる。

**補題 4.2** $(p,p_w)$ を $w$ の最分割汎化対とする。このとき，包含関係 $L(p,p_w) \subset L(q_1,q_2)$ を満たす $(q_1,q_2) \in P_{w,X}$ は存在しない。

(証明) 包含関係 $L(p,p_w) \subset L(q_1,q_2)$ を満たす $(q_1,q_2) \in P_{w,X}$ が存在すると仮定する。補題 3.1

より，mi $\{p, p_w\} = Q_{w,d_p}$ である。仮定より，任意の $\xi \in Q_{w,p_w}$ に対して，$\xi \preceq \eta$ となる長さ $n+1$ の $q_1, q_2$ の極大例化 $\eta$ が存在しなければならない。$(p, p_w)$ は，$w$ の最分割汎化対であるので，$d_p \cap d_{p_w} = \{2, 3, \cdots, n\}$ である。従って，位置 $l > 1$ に変数をもつ $\xi \in Q_{w,d_p}$ が存在する。補題 3.3(1) より，$l$ が $q_1, q_2$ の $c$ 連続点ではないことを意味する。すなわち，$q_1, q_2$ ともに，定数と変数が交互に現れる。smi $\{q_1, q_2\} = w$ より，このような対は，$(p, p_w)$ 以外にはない。これは，仮定に反する。■

$r \preceq s$ となるパターン $r, s$ の分割集合を $d_{r,s} = \{l_1, l_2, \cdots, l_k\}$ とする。$r, s$ に対して，次の集合を導入する:

$$I_{r,s} = \{l_i + j_i \mid l_i \in \hat{d}_{r,s},\ 0 < j_i < l_{i+1} - l_i\},$$
$$\hat{P}_{r,s} = \{(p, q) \in P_{r,s} \mid l_i \in \hat{d}_{r,s} \text{ に対して，} (p_i, q_i) \text{ は } w_i \text{ の最分割汎化対となる } (i = 1, 2, \cdots, k+1)\}$$

ただし，上記の $p_i, q_i$ は，$p, q$ の $d_{r,s}$-表示における $i$ 番目の部分列を表すものとする。次の結果は，$I_{r,s}$ の定義から直ちに示される。

**補題 4.3** $(p, q) \in \hat{P}_{r,s}, d_{r,s} = \{l_1, \cdots, l_k\}$ とする。このとき，$I_{r,s} \cap d_{r,s} = \phi$ かつ $d_{p,q} = d_{r,s} \cup I_{r,s}$ である。

上記の結果より，$d_{p,q}$ は，$(p, q) \in \hat{P}_{r,s}$ のとり方によらず，$r, s$ のみで定まることに注意する。

**補題 4.4** $(p, q) \in P_{r,s}$ とする。$I_{r,s} \subseteq d_p \cap d_q$ ならば，$(p, q) \in \hat{P}_{r,s}$ である。

(証明) $d_{r,s} = \{l_1, \cdots, l_k\}$ とし，$p, q$ の $d_{r,s}$-表現を $p = p_1 \cdots p_{k+1}, q = q_1 \cdots q_{k+1}$ とする。$l = l_i + j_i \in I_{r,s}$ とする。ただし，$l_i \in \hat{d}_{r,s}, 0 < j_i < l_{i+1} - l_i$ とする。$I_{r,s}$ の定義より，$l$ は，$r$ の $c$ 連続点である。仮定より，$l \in d_p \cap d_q$ であるが，$l$ が $p, q$ の $cv$ 切断点 ($vc$ 切断点) で，$p[l], q[l] \in X$ ならば，$r[l] \in X$ となり，$r[l] \in \Sigma$ であることに反する。従って，$l$ が $p$ の $cv$ 切断点ならば，$q$ に対しては $vc$ 切断点でなければならない。任意の $l \in I_{r,s}$ に対して，これが成り立つ。これは，$(p, q) \in \hat{P}_{r,s}$ であることを意味する。■

**定理 4.5** $(p_1, p_2) \in \hat{P}_{r,s}$, $(q_1, q_2) \in P_{r,s}$ とする。このとき，$L(p_1, p_2) \subseteq L(q_1, q_2)$ ならば，$(q_1, q_2) \in \hat{P}_{r,s}$ である。

(証明) $(p_1, p_2) \in \hat{P}_{r,s}$, $(q_1, q_2) \in P_{r,s}$, $L(p_1, p_2) \subseteq L(q_1, q_2)$ とする。$\hat{P}_{r,s} \subseteq P_{r,s}$ より，$d_{p_1} \cap d_{p_2} = I_{r,s}$ が成り立つので，定理4.1(1) より，$d_{p_1} \cap d_{p_1} \subseteq d_{q_1,q_2}$ である。補題3.3(2) より，各 $l \in d_{p_1} \cap d_{p_2}$ に対して，位置 $l$ が変数となるパターン $\xi \in$mi $\{p_1, p_2\}_{n+1}$ が存在する。共通言語の包含関係から，$\xi$ は，$q_1, q_2$ の例化である。$l \in d_{q_1,q_2}$ であるので，補題3.3(1) より，このような $l$ は，$q_1, q_2$ の $c$ 連続点にはならない。$(q_1, q_2) \in P_{r,s}, r[l] \in \Sigma$ であるので，$l \in d_{q_1} \cap d_{q_2}$ でなければならない。すなわち，$I_{r,s} \subseteq d_{q_1} \cap d_{q_2}$ である。これは，補題4.4 より，$(q_1, q_2) \in \hat{P}_{r,s}$ であることを意味する。■

## 参考文献


[1] D. Angluin, *Finding patterns common to a set of strings*, in Proceedings of the 11th Annual Symposium on Theory of Computing, (1979) 130–141.

[2] Y. Mukouchi, *Containment problems for pattern languages*, IEICE Transactions on Information and Systems, **E75-D**(4) (1992) 260–267.

[3] M. Sato and Y. Mukouchi, *Learning of Languages Generated byPatterns from Positive Examples*, in Proceedings of BIOCOMP2002, (2002), 22.

[4] T. Shinohara, *Polynomial time inference of pattern languages and its applications*, in Preceedings of the 7th IBM Symposium on Mathematical Foundations of Computer Science, (1982) 191–209.

[5] 寺田幹治，向内康人，佐藤優子，正例からのパターン上の決定木の帰納推論，電子情報通信学会論文誌，**J83-D-I**(1) (2000) 60–67.

[6] Jin Uemura and Masako Sato, *Compactness and Learning of Classes of Unions of Erasing Regular Pattern Language*, Algoruthmic Learning Theory in Proceedings of the 13th International Conference, ALT 2002 (2002) 293-307.