正規パターン言語の有限和に対する特徴集合とコンパクト性について

武田 直人 * 内田 智之 * 正代 隆義 †
松本 哲志 ‡ 鈴木 祐介 * 宮原 哲浩 *

概要

正規パターンとは，定数記号と変数記号からなる，各変数記号が高々1 回しか出現しない記号列をいう．正規パターン p の変数記号を定数記号列で置き換えることで生成できる定数記号列全体の集合を $L(p)$ で表す．高々k $(k \geq 3)$ 個の正規パターンの集合全体のクラスを $\mathcal{RP}^k$ で表す．1998 年に Sato ら [1] は，各変数記号に対し長さが高々2 の記号列を代入することで $P \in \mathcal{RP}^k$ から得られる記号列の有限集合 $S_2(P)$ が，$L(P) = \bigcup_{p \in P} L(p)$ の特徴集合であることを示した．つぎに，定数記号の数が $2k-1$ 以上のとき，$\mathcal{RP}^k$ が包含に関してコンパクト性をもつことを示した．これらの結果に対し，本稿では，まず Sato ら [1] の結果を検証し，Sato らが与えた定理の証明の誤りを修正した．さらに，隣接した変数（隣接変数）を持たない正規パターンである非隣接変数正規パターン全体の集合 $\mathcal{RP}_{\mathcal{NAV}}$ を与え，高々k $(k \geq 1)$ 個の非隣接変数正規パターンの集合全体のクラス $\mathcal{RP}^k_{\mathcal{NAV}}$ に属する集合 P から得られる $S_2(P)$ が $L(P)$ の特徴集合であることを示した．さらに，定数記号の数が $k + \sqrt{k+1}$ 以上のとき，$\mathcal{RP}^k_{\mathcal{NAV}}$ が包含に関してコンパクト性をもつことを示した．これにより，正規パターン言語のときよりも少ない定数記号の数で，非隣接変数正規パターン言語の有限和に関する効率的な学習アルゴリズムが設計できることを示した．

*広島市立大学大学院情報科学研究科知能工学専攻
(mh67011@e.hiroshima-cu.ac.jp)
†福岡工業大学情報工学部情報工学科
‡東海大学理学部情報数理学科

1 はじめに

パターンとは，定数記号と変数記号からなる記号列である．例えば，a, b, c を定数記号，x, y を変数記号とするとき，$axbxcy$ はパターンである．パターン p に対し，すべての変数記号を空記号列 ε でない定数記号列で置き換えて得られる記号列の集合を，p により生成される**パターン言語**あるいは単にパターン言語といい，$L(p)$ と書く．なお，同じ変数記号には同じ定数記号列で置き換える．例えば，上記のパターン $axbxcy$ により生成されるパターン言語 $L(axbxcy)$ は $\{aubucw \mid w, u$ は ε でない定数記号列 $\}$ である．各変数記号が高々1 回しか現れないパターンを正規パターンという．例えば，パターン $axbxcy$ は正規パターンではないが，変数記号 x, y, z を持つパターン $axbzcy$ は正規パターンである．パターン q がパターン p の変数記号をパターンで置き換えることで得られるとき，q は p の汎化といい，$p \preceq q$ と書く．例えば，パターン $q = axz$ はパターン $p = axbxcy$ の汎化である．なぜならば，q の変数記号 z をパターン $bxcy$ で置き換えると p が得られるからである．よって，$p \preceq q$ である．パターン p, q に対して，もし $p \preceq q$ ならば，$L(p) \subseteq L(q)$ となることは明らかである．しかし，その逆，つまり $L(p) \subseteq L(q)$ ならば $p \preceq q$ は成り立つとは限らない．これに対し，Mukouchi[2] は，任意の正規パターン p, q については，$L(p) \subseteq L(q)$ ならば $p \preceq q$ も成り立つことを示した．

$\mathcal{RP}^k$ を高々k $(k \geq 3)$ 個の正規パターンの集合全体のクラスとする．正規パターンの集合 $P \in \mathcal{RP}^k$ に対し，$L(P) = \bigcup_{p \in P} L(p)$ とし，$\mathcal{RP}^k$ に対する正規

パターン言語のクラス $\mathcal{RPL}^k$ を $\{L(P) \mid P \in \mathcal{RP}^k\}$ とする．$P, Q \in \mathcal{RP}^k$ とする．任意の正規パターン $p \in P$ に対し，正規パターン $q \in Q$ が存在し，$p \preceq q$ が成り立つとき $P \sqsubseteq Q$ と書く．定義より，$P \subseteq Q$ ならば $L(P) \subseteq L(Q)$ であることは明らかである．そこで，Sato ら [1] は，$k \leq 3$ であり定数記号の数が $2k-1$ であるとき，各変数記号に対し長さが高々2 の記号列を代入することで $P \in \mathcal{RP}^k$ から得られる記号列の有限集合 $S_2(P)$ が $L(P)$ の特徴集合であること，つまり任意の正規パターン言語 $L' \in \mathcal{RPL}^k$ に対して，$S_2(P) \subseteq L'$ ならば $L(P) \subseteq L'$ となることを示すで，$(i) S_2(P) \subseteq L(Q)$，(ii) $P \sqsubseteq Q$ および (iii) $L(P) \subseteq L(Q)$ が同値であることを示した．しかし，この結果の根拠となる補題 14[1] に誤りがあるため，本稿では，まずその修正を行い，Sato らが示した 3 つの命題の同値性の正しい証明を与えた．Sato ら [1] は，定数記号の数が $2k-1$ 以上のとき，$\mathcal{RP}^k$ が包含に関してコンパクト性をもつことも示した．これに対し，隣接した変数（隣接変数）を持たない正規パターンである非隣接変数正規パターン全体の集合 $\mathcal{RP}_{\mathcal{NAV}}$ を与え，高々k $(k \geq 1)$ 個の非隣接変数正規パターンの集合全体のクラス $\mathcal{RP}^k_{\mathcal{NAV}}$ に属する集合 P から得られる $S_2(P)$ が $L(P)$ の特徴集合であることを示した．さらに，定数記号の数が $k+\sqrt{k+1}$ 以上のとき，$\mathcal{RP}^k_{\mathcal{NAV}}$ が包含に関してコンパクト性をもつことを示した．これは，正規パターン言語のときに下界を示すために用いられた関数 $2k-1$ より正確な関数表現 $k+\sqrt{k+1}$ を与えることができ，その定数記号の数の条件で，非隣接変数正規パターン言語の有限和に関する効率的な学習アルゴリズムが設計できることを示した．

本稿の構成は以下の通りである．第 2 節では，準備としてパターン言語，正規パターン言語，コンパクト性などの定義を与え，さらに $\mathcal{RP}^+$ の特徴集合に関する Sato らの結果を紹介する．第 3 節では，$S_2(P)$ は $\mathcal{RPL}^k$ における $L(P)$ の特徴集合であること，および $\mathcal{RP}^k$ が包含に関するコンパクト性を持つことを示す．第 4 節では，非隣接変数正規パターンを与え，$\mathcal{RP}^k_{\mathcal{NAV}}$ に属する集合 P から得られる $S_2(P)$ が $L(P)$ の特徴集合であること，および $\mathcal{RP}^k_{\mathcal{NAV}}$ が包含に関してコンパクト性をもつことを示す．

2　準備

Σ を有限アルファベットとし，X を $\Sigma \cap X = \emptyset$ を満たす加算無限集合とする．Σ と X の要素をそれぞれ定数記号と変数記号という．Σ と X の記号からなる記号列を**パターン**という．また，各変数記号が高々 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，つまりその記号列の長さを $|p|$ で表す．すべてのパターンの集合とすべての正規パターンの集合をそれぞれ $\mathcal{P}$ と $\mathcal{RP}$ で表す．便宜上，空記号列 ε もパターンとしていることに注意する．つまり，$\mathcal{P} = (\Sigma \cup X)^*$ であり，$\mathcal{P} \setminus \{\varepsilon\} = (\Sigma \cup X)^+$ である．集合 A の要素数を $\sharp A$ で表す．本稿では，$\sharp\Sigma \geq 2$ と仮定する．$\mathcal{P}$ の要素を $p, q, \cdots, p_1, p_2, \cdots,$ で表す．

与えられたパターンの変数に長さ 1 以上のパターンを代入することで，別のパターンを生成することができる．ただし，同じ変数記号には同じパターンを代入し，空記号列 ε は代入しないこととする．パターン $p \in \mathcal{P}$ に対し，p 中の各変数 x_i $(i = 1, 2, \ldots, k)$ にそれぞれパターン q_i を代入することを $\theta = \{x_1 := q_1, x_2 := q_2, \cdots, x_k := q_k\}$ で表すこととし，このような代入の操作を p に施した結果のパターンを $p\theta$ で表す．便宜上 θ を代入と呼ぶ．q が p の汎化，あるいは p が q の例化であるとは，$p = q\theta$ を満たす代入 θ が存在するときをいい，$\boldsymbol{p \preceq q}$ で表す．また，$p \preceq q$ かつ $q \preceq p$ であるとき，p と q は等価であるといい，$\boldsymbol{p \equiv q}$ で表す．

パターン p に対し，p が表す言語 (Σ^* の部分集合) を，p に代入を施すことにより生成できる定数記号列の集合 $L(p)$，つまり，$L(p) = \{w \in \Sigma^+ \mid w \preceq p\}$ と定義する．ここで，$p \equiv q$ ならば $L(p) = L(q)$ であることに注意する．パターンおよび正規パターンによって生成される言語をそれぞれパターン言語および正規パターン言語という．また，すべてのパターン言語の集合および正規パターン言語の集合をそれぞれ $\mathcal{PL}$ およびび $\mathcal{RPL}$ で表す．正規パターンにつ

いては，次の補題が成り立つ．

補題 1 (Mukouchi[3])**.** p, q を正規パターンとする．このとき，$p \preceq q$ ならば $L(p) \subseteq L(q)$ である．

なお，一般に逆は成り立たないことに注意する．

$\mathcal{P}$ の空でない部分集合を $\mathcal{P}^+$ で，高々k $(k \geq 1)$ 個のパターンからなる $\mathcal{P}$ の部分集合を $\mathcal{P}^k = \{P \in \mathcal{P}^+ \mid |P| \leq k\}$ で表す．また，高々k $(k \geq 1)$ 個のパターン集合 $P \in \mathcal{P}^k$ に対して，P が表すパターン言語 $L(P)$ を $L(P) = \bigcup_{p \in P} L(p)$ で，$\mathcal{P}^k$ に属するパターン集合が表すパターン言語のクラス $\mathcal{PL}^k$ を $\mathcal{PL}^k = \{L(P) \mid P \in \mathcal{P}^k\}$ で表す．同様に，$\mathcal{RP}$ の空でない部分集合を $\mathcal{RP}^+$ で，高々k $(k \geq 1)$ 個のパターンからなる $\mathcal{RP}$ の部分集合を $\mathcal{RP}^k = \{P \in \mathcal{RP}^+ \mid |P| \leq k\}$, $\mathcal{RP}^k$ に属するパターン集合が表すパターン言語のクラスを $\mathcal{RPL}^k = \{L(P) \mid P \in \mathcal{RP}^k\}$ で表す．P, Q を $\mathcal{P}^+$ に属するパターンとする．このとき，任意のパターン $p \in P$ に対して，あるパターン $q \in Q$ が存在し，$p \preceq q$ が成り立つとき $P \sqsubseteq Q$ と書く．

補題 2 (Mukouchi[3])**.** P, Q を $\mathcal{P}^+$ に属するパターンとする．このとき，$P \sqsubseteq Q$ ならば $L(P) \subseteq L(Q)$ である．

なお，一般にこの逆は成り立たないことに注意する．

定義 3. クラス $\mathcal{C} \subseteq \mathcal{P}^+$ が**包含に関するコンパクト性を持つ**とは，任意のパターン $p \in \mathcal{P}$ と任意のパターン集合 $Q \in \mathcal{C}$ に対して，$L(p) \subseteq L(Q)$ ならば，ある $q \in Q$ が存在して $L(p) \subseteq L(q)$ であるときをいう．

同様にして，クラス $\mathcal{C} \in \mathcal{RP}^+$ が包含に関するコンパクト性を持つことが定義できる．補題 1 より，全ての $P, Q \in \mathcal{RP}^+$ に対して，$P \sqsubseteq Q$ ならばそのときに限り $L(P) \subseteq L(Q)$ であることが示せる．さらに，$\mathcal{RP}^k$ については，任意のパターン $p \in \mathcal{C}$ に対し，ある特定の有限部分集合 $S \subseteq L(p)$ が存在して，$S \subseteq L(Q)$ ならば，ある $q \in Q$ に対して $L(p) \subseteq L(q)$ となることが知られている [3]．また，$S \subseteq L(Q)$ ならば $L(p) \subseteq L(Q)$ である．これにより，S は次の定義される $L(p)$ の特徴集合であることがわかる．

定義 4. $\mathcal{L}$ を言語クラスとする．L を $\mathcal{L}$ に属する言語とする．空でない有限部分集合 $S \subseteq \Sigma^+$ は $\mathcal{L}$ における L の**特徴集合**であるとは，任意の $L' \in \mathcal{L}$ に対して $S \subseteq L'$ ならば $L \subseteq L'$ となるときをいう．

この特徴集合の概念は，有限弾力性 [4] と有限交差性 [**?**] に密接な関係があることが知られている．

m $(m \geq 0)$ 個の変数記号 $x_1, \ldots, x_m$ を含む正規パターン p と n $(n \geq 1)$ に対して，p 中の各変数記号に長さが高々n の Σ^+ の定数記号列を代入して得られるすべての定数記号列の集合を $S_n(p)$ で表す．さらに，正規パターンの空でない有限集合 P に対して，$S_n(P) = \bigcup_{p \in P} S_n(p)$ とする．このとき，任意の n $(n \geq 1)$ に対して，$S_n(P) \subseteq S_{n+1}(P) \subseteq L(P)$ である．よって，次の定理が成り立つ．

定理 5 (Sato et al.[1])**.** 任意の $P \in \mathcal{RP}^k$ に対して，$S_n(P)$ がクラス $\mathcal{RPL}^k$ 内の正規パターン言語 $L(P)$ の特徴集合であるような n $(n \geq 1)$ が存在する．

p_1, p_2, r, q を正規パターンとし，$p_1 r p_2 \preceq q$ が成り立つとする．また，$x_1, \ldots, x_n$ を q に含まれる変数記号とする．このとき，$q = q_1 x_i q_2$ に対して，$p_1 = (q_1\theta)r'$ かつ $p_2 = r''(q_2\theta)$ を満たす変数記号 x_i と代入 $\theta = \{x_1 := r_1, \ldots, x_i := r' r r'', \ldots, x_n := r_n\}$ が存在すれば，$p_1 r p_2$ に含まれる部分パターン r は q の変数記号への代入により生成できる．よって，$p_1 r p_2$ に含まれる部分パターン r が q の変数記号への代入により生成できるとき，$p_1 x p_2 \preceq q$ が成り立つ．

補題 6 (Sato et al.[1])**.** $p = p_1 x p_2$, $q = q_1 q_2 q_3$ を正規パターンとする．以下の (i), (ii), (iii) がすべて成り立つとき，$p \preceq q$ である．

(i) $p_1 \preceq q_1 q_2$,
(ii) $p_2 \preceq q_2 q_3$,
(iii) q_2 は変数を含む.

ある $a \in \Sigma$ に対して，$p\{x := a\} \preceq q$ のとき，$p_1 x p_2 \not\preceq q$ ならば，$p_1 a p_2$ の定数記号 a は，q の変数記号への代入によって生成することはできない．すなわち，$p_1 \preceq q_1$ かつ $p_2 \preceq q_2$ を満たす $q = q_1 a q_2$ が存在する．これにより，次の補題が得られる．

補題 7 (Sato et al.[1]). $\sharp\Sigma \geq 3$, $p = p_1xp_2$, q を正規パターン, a, b, c を Σ に属する相異なる定数記号とする．このとき，$p_1ap_2 \preceq q$, $p_1bp_2 \preceq q$ かつ $p_1cp_2 \preceq q$ が成り立つならば，$p \preceq q$ が成り立つ．

次の補題 8 は，相異なる定数記号 a, b に対して，$p\{x := a\} \preceq q$ かつ $p\{x := b\} \preceq q$ ならば $p \not\preceq q$ となる正規パターン p, q が存在することを示している

補題 8 (Sato et al.[1]). $\sharp\Sigma \geq 3$ とする．$a, b \in \Sigma$ を相異なる定数記号とする．次の条件 (i), (ii) を満たす正規パターン $p = p_1AwxwBp_2$ と $q = q_1AwBq_2$ に対して，$p\{x := a\} \preceq q$ かつ $p\{x := b\} \preceq q$ ならば $p \not\preceq q$ である．ここで，p_1, p_2, q_1, q_2 は正規パターン，w は定数記号列である．

(i) $p_1 \preceq q_1$, (ii) $p_2 \preceq q_2$,
(iii) $A = a, B = b$ または $A = b, B = a$.

補題 7 より，次の定理が成り立つ．

定理 9 (Sato et al.[1]). $\sharp\Sigma \geq 2k + 1$ とし，$P \in \mathcal{RP}^+$, $Q \in \mathcal{RP}^k$ とする．このとき，次の (i), (ii), (iii) は同値である．

(i) $S_1(P) \subseteq L(Q)$,
(ii) $P \sqsubseteq Q$,
(iii) $L(P) \subseteq L(Q)$.

この定理により，次の系が得られる．

系 10 (Sato et al.[1]). $\sharp\Sigma \geq 3$ とし，p, q を正規パターンとする．このとき，次の (i), (ii), (iii) は同値である．

(i) $S_1(p) \subseteq L(q)$,
(ii) $p \preceq q$,
(iii) $L(p) \subseteq L(q)$.

3 $S_2(P)$ とコンパクト性

この節では，$\sharp\Sigma \geq 2k - 1$ と仮定したとき，$S_2(P)$ は $\mathcal{RPL}^k$ における $L(P)$ の特徴集合であることを示し，$\mathcal{RP}^k$ が包含に関するコンパクト性を持つことを示す．

補題 11. $\sharp\Sigma \geq 4$ とし，p, q を正規パターンとする．正規パターンの有限集合 D が，次の (i-1), (i-2), (ii) のいずれかで表されるとき，すべての $r \in D$ に対して $p\{x := r\} \preceq q$ ならば，$p\{x := xy\} \preceq q$ である．

(i-1) $\{ay, by\}$ $(a \neq b)$,

(i-2) $\{ya, yb\}$ $(a \neq b)$,

(ii) $\{a_1b_1, a_2b_2, a_3b_3, a_4b_4\}$ ($i \neq j$ に対して，$a_i \neq a_j$ かつ $b_i \neq b_j$).

証明. p に変数記号が含まれない場合は自明である．したがって，正規パターン p には変数記号が現れるとし，その変数記号を x とする．このとき，正規パターン p_1, p_2 が存在し，$p = p_1xp_2$ と表すことができる．$p\{x := xy\} \not\preceq q$ と仮定して，矛盾を導く．

(i-1) $D = \{ay, by\}$ $(a \neq b)$ であるとする．$p\{x := xy\} \preceq q$ ではなく，$p_1ayp_2 \preceq q$ かつ $p_1byp_2 \preceq q$ であることから，正規パターン q_1, q_2 と変数記号 y_1, y_2, さらに定数記号列 w が存在して，$q = q_1ay_1wby_2q_2$ または $q = q_1by_1way_2q_2$ と表すことができる．$q = q_1ay_1wby_2q_2$ と表されるとき，次の (1), (2), (1'), (2') が全て成り立つ．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq wby_2q_2$ または $p_2 \preceq y'wby_2q_2$
(2)	$p_1 \preceq q_1ay_1$	(2')	$p_2 \preceq q_2$ または $p_2 \preceq y''q_2$

(2) より，正規パターン p_1', p_1'' が存在して，

$$p_1 = p_1'p_1'',\ p_1' \preceq q_1a,\ p_1'' \preceq y_1w$$

が成り立つ．したがって，$p = p_1xp_2 = p_1'p_1''xp_2$ であるから，(1') が $p_2 \preceq wby_2q_2$ のとき，$p \preceq q_1ap_1''xwby_2q_2 = q\{y_1 := p_1''x\}$ となる．また，(1') が $p_2 \preceq y'wby_2q_2$ のとき，$p \preceq q_1ap_1''xy'wby_2q_2 = q\{y_1 := p_1''xy'\}$ となる．よって，$p \preceq q$ が成り立ち，仮定 $p\{x := xy\} \not\preceq q$ に矛盾する．

(i-2) $D = \{ya, yb\}$ $(a \neq b)$ のときは，記号列 p と q を逆順にすることにより，**(i-1)** の場合と同様に，仮定 $p\{x := xy\} \not\preceq q$ に矛盾することを証明できる．

(ii) $D=\{a_1b_1,a_2b_2,a_3b_3,a_4b_4\}$ ($i\neq j$ に対して，$a_i\neq a_j$ かつ $b_i\neq b_j$) であるとする．すべての $r\in D$ に対して $p\{x:=r\}\preceq q$ であることから，正規パターン q には，$a_1b_1,a_2b_2,a_3b_3,a_4b_4$ に対応する 4 つの長さ 2 の記号列が存在する．その 4 つの記号列は一部を重複して現れることがあることに注意する．D の 4 つの記号列に対応する q の記号列の現れ方には次の 15 通り存在する．

(a) $a_1b_1,a_2b_2,a_3b_3,a_4b_4$	(i) a_1b_1,yb_2,a_3y,a_4y
(b) $a_1b_1,a_2b_2,a_3b_3,a_4y$	(j) a_1b_1,a_2y,a_3y,a_4y
(c) $a_1b_1,a_2b_2,a_3b_3,yb_4$	(k) yb_1,yb_2,yb_3,yb_4
(d) a_1b_1,a_2b_2,a_3y,yb_4	(l) yb_1,yb_2,yb_3,a_4y
(e) a_1b_1,a_2b_2,yb_3,yb_4	(m) yb_1,yb_2,a_3y,a_4y
(f) a_1b_1,a_2b_2,a_3y,a_4y	(n) yb_1,a_2y,a_3y,a_4y
(g) a_1b_1,yb_2,yb_3,yb_4	(o) a_1y,a_2y,a_3y,a_4y
(h) a_1b_1,yb_2,yb_3,a_4y	

上記 (e)–(o) の 11 通りの記号列を含む正規パターン q は，補題 11 の (i-1) または (i-2) に対応する記号列が現れる．したがって，その場合の証明より仮定 $p\{x:=xy\}\not\preceq q$ に矛盾する．したがって，(a)–(d) の 4 通りついて矛盾を導く．

(a), (b), (c) は，q に a_1b_1,a_2b_2,a_3b_3 が現れる場合，(d) は，q に a_1b_1,a_2b_2,a_3y が現れる場合と q に a_1b_1,a_2b_2,yb_4 が現れる場合である．本論文では，q に a_1b_1,a_2b_2,a_3y が現れる場合を証明する．q に a_1b_1,a_2b_2,yb_4 が現れる場合は，記号列 p と q を逆順にすることにより，q に a_1b_1,a_2b_2,a_3y が現れる場合の証明から導かれる．

(abc) q に a_1b_1,a_2b_2,a_3b_3 が現れる場合: このとき，それら 3 つの記号列が重複する場合があるので，次の 3 つの場合に分けて証明する．

(abc-1) $q=q_1a_1b_1wa_2b_2w'a_3b_3q_2$,

(abc-2) $q=q_1a_1b_1a_3b_3q_2$ $(b_1=a_2,a_3=b_2)$,

(abc-3) $q=q_1a_1b_1b_2wa_3b_3q_2$ $(b_1=a_2)$.

(abc-1) $q=q_1a_1b_1wa_2b_2w'a_3b_3q_2$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1\preceq q_1$	(1')	$p_2\preceq wa_2b_2w'a_3b_3q_2$
(2)	$p_1\preceq q_1a_1b_1w$	(2')	$p_2\preceq w'a_3b_3q_2$
(3)	$p_1\preceq q_1a_1b_1wa_2b_2w'$	(3')	$p_2\preceq q_2$

$|w|=|w'|$ であれば，(2) と (3) より，p_1 の接尾辞は $a_1b_1wa_2b_2w'$ かつ a_1b_1w であるので，$a_1b_1w=a_2b_2w'$ となる．よって，$a_1b_1=a_2b_2$ となり，$a_i\neq a_j$ かつ $b_i\neq b_j(i\neq j)$ であることに矛盾する．

$|w|+1=|w'|$ であれば，(1') と (2') より，p_2 の接頭辞は $wa_2b_2w'a_3b_3$ かつ $w'a_3b_3$ である．$w'=ww_1$ とおくと，$w'a_3b_3=ww_1a_3b_3$ となる．したがって，$wa_2b_2=ww_1a_3$ より $b_2=a_3$ となる．(2) と (3) より，p_1 の接尾辞は $a_1b_1wa_2b_2w',a_1b_1w$ である．$w'=w_2w$ とおくと，$a_1b_1wa_2b_2w'=a_1b_1wa_2b_2w_2w$ となる．したがって，$b_2w_2w=a_1b_1w$ より，$b_2=a_1$ となる．$b_2=a_3$ より，$a_3=a_1$ となり，$a_i\neq a_j$ $(i\neq j)$ であることに矛盾する．

$|w|+1<|w'|$ であれば，(2) と (3) より，p_1 の接尾辞は $a_1b_1wa_2b_2w'$ かつ a_1b_1w である．$w'=w_1w$ とおくと，$a_1b_1wa_2b_2w'=a_1b_1wa_2b_2w_1w$ となる．$|w_1|\geq 2$ であるため，w_1 の接尾辞は a_2b_2 となる．$(1'),(2')$ より，p_2 の接頭辞は $wa_2b_2w'a_3b_3$ かつ $w'a_3b_3$ である．$w'=ww_2$ とおくと，$w'a_3b_3=ww_2a_3b_3$,$w'=w_1w$ とおくと，$wa_2b_2w'a_3b_3=wa_2b_2w_1wa_3b_3$ となる．$|ww_2a_2b_2|=|wa_2b_2w_1|$ より，w_1 の接尾辞は a_3b_3 となる．よって，w_1 の接尾辞は $a_2b_2=a_3b_3$ となり，$a_i\neq a_j$ かつ $b_i\neq b_j(i\neq j)$ であることに矛盾する．

(abc-2) $q=q_1a_1b_1a_3b_3q_2$ $(b_1=a_2,a_3=b_2)$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1\preceq q_1$	(1')	$p_2\preceq a_3b_3q_2$
(2)	$p_1\preceq q_1a_1$	(2')	$p_2\preceq b_3q_2$
(3)	$p_1\preceq q_1a_1b_1$	(3')	$p_2\preceq q_2$

(2) と (3) より，p_1 の接尾辞は a_1b_1 かつ a_1 であり，$b_1=a_1$ となる．$b_1=a_2$ より，$a_1=a_2$ であるため, $a_i\neq a_j$ $(i\neq j)$ であることに矛盾する．

(abc-3) $q=q_1a_1b_1b_2wa_3b_3q_2$ $(b_1=a_2)$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq b_2wa_3b_3q_2$
(2)	$p_1 \preceq q_1a_1$	(2')	$p_2 \preceq wa_3b_3q_2$
(3)	$p_1 \preceq q_1a_1b_1b_2w$	(3')	$p_2 \preceq q_2$

$w = \varepsilon$ のとき，(2) と (3) より，p_1 の接尾辞は a_1 かつ $a_1b_1b_2$ であり，(1') と (2') より，p_2 の接頭辞は $b_2a_3b_3$ かつ a_3b_3 である．$b_2 = a_1$ と $b_2a_3 = a_3b_3$ より，$a_1 = a_3$ となり，$a_i \neq a_j$ $(i \neq j)$ であることに矛盾する．$|w| \geq 1$ のとき，(2) と (3) より，p_1 の接尾辞は a_1 かつ $a_1b_1b_2w$ である．よって，w の接尾辞は a_1 となる．(1') と (2') より，p_2 の接頭辞は $b_2wa_3b_3$ かつ wa_3b_3 となる．よって，w の接尾辞は a_3 となる．したがって，w の接尾辞は $a_1 = a_3$ となり，$a_i \neq a_j$ $(i \neq j)$ であることに矛盾する．

(d) q に a_1b_1, a_2b_2, a_3y が現れる場合: このとき，記号列 A, B, C に対して，$\{A, B, C\} = \{a_1b_1, a_2b_2, a_3y\}$ とおき，$q = q_1AwBw'Cq_2$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq wBw'Cq_2$
(2)	$p_1 \preceq q_1Aw$	(2')	$p_2 \preceq w'Cq_2$
(3)	$p_1 \preceq q_1AwBw'$	(3')	$p_2 \preceq q_2$

$|w| = |w'|$ であれば，(2) と (3) より，p_1 の接尾辞は Aw かつ $AwBw'$ である．よって，$Aw = Bw'$ となり，$A \neq B$ であることに矛盾する．

$|w| \neq |w'|$ とする．$A = a_3y$ とすると，$B = a_1b_1$，$C = a_2b_2$ としてよいので，(2) は $p_1 \preceq q_1a_3yw$ となる．したがって，正規パターン p'_1, p''_1 が存在して，$p_1 = p'_1p''_1$，$p'_1 \preceq q_1a_3$ かつ $p''_1 \preceq yw$ となる．これらと (1') より，$p = p_1xp_2 = p'_1p''_1xp_2 \preceq q_1a_3p''_1xwBw'Cq_2 = q\{y := p''_1x\}$ となり，$p = q\theta$ となる．これは仮定に矛盾する．$B = a_3y$ とすると，$A = a_1b_1$，$C = a_2b_2$ としてよいので，(3) は $p_1 \preceq q_1a_1b_1wa_3yw'$ となり，(1') は $p_2 \preceq wa_3yw'a_2b_2q_2$ である．$q'_1 = q_1a_1b_1$，$q'_2 = wa_3yw'$，$q'_3 = a_2b_2q_2$ とおくと，$(3)p_1 \preceq q'_1q'_2, (1')p_2 \preceq q'_2q'_3, q'_2$ は変数記号が含まれる．補題 6 より，$p \preceq q$ となり，$p\{x := xy\} \preceq q$ である．これは仮定に矛盾する．

以上より，A または B が a_3y の場合，仮定に矛盾するため，$C = a_3y$ となる．$C = a_3y$ のとき，3 つの記号列が重複する場合を考慮して，次の 5 つの場合に分けて証明する．

(d-1) $q = q_1a_1b_1wa_2b_2w'a_3yq_2$,

(d-2) $q = q_1a_1b_1b_2yq_2$ $(a_2 = b_1, a_3 = b_2)$,

(d-3) $q = q_1a_1b_1b_2wa_3yq_2$ $(b_1 = a_2)$,

(d-4) $q = q_1a_3ywa_1b_1b_2q_2$ $(b_1 = a_2)$,

(d-5) $q = q_1a_1b_1ywa_2b_2q_2$ $(b_1 = a_3)$.

(d-1) $q = q_1a_1b_1wa_2b_2w'a_3yq_2$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq wa_2b_2w'a_3yq_2$
(2)	$p_1 \preceq q_1a_1b_1w$	(2')	$p_2 \preceq w'a_3yq_2$
(3)	$p_1 \preceq q_1a_1b_1wa_2b_2w'$	(3')	$p_2 \preceq q_2$

$|w|+1 = |w'|$ のとき，(2) と (3) より，p_1 の接尾辞は $a_1b_1wa_2b_2w'$ かつ a_1b_1w である．$w' = w_1w$ とおくと，$a_1b_1wa_2b_2w' = a_1b_1wa_2b_2w_1w$ と表すことができる．$b_2w_1w = a_1b_1w$ より，$b_2 = a_1$ となる．(1') と (2') より，p_2 の接頭辞は $wa_2b_2w'a_3$ かつ $w'a_3$ である．$w' = ww_2$ とおくと，$w'a_3 = ww_2a_3$ と表すことができる．$wa_2b_2 = ww_2a_3$ より，$b_2 = a_3$ となる．よって，$b_2 = a_1$ より，$a_1 = a_3$ となり，$a_i \neq a_j$ $(i \neq j)$ であることに矛盾する．

$|w'|+1 = |w|$ のとき，(1') と (2') より，p_2 の接頭辞は $wa_2b_2w'a_3$ かつ $w'a_3$ である．$w = w'w_1$ とおくと，$wa_2b_2w'a_3 = w'w_1a_2b_2w'a_3$ と表すことができる．$w'w_1 = w'a_3$ より，$w_1 = a_3$ となる．(2) と (3) より，p_1 の接尾辞は $a_1b_1wa_2b_2w'$ かつ a_1b_1w である．$w = w'w_1$ とおくと，$a_1b_1wa_2b_2w' = a_1b_1w'w_1a_2b_2w'$ となる．さらに，$w = w_2w'$ とおくと，$a_1b_1w = a_1b_1w_2w'$ と表すことができる．$|a_2b_2w'| = |b_1w_2w'|$ より，$w_1 = a_1$ となる．よって，$w_1 = a_3$ より，$a_1 = a_3$ となり，$a_i \neq a_j$ $(i \neq j)$ であることに矛盾する．

$|w|+1 < |w'|$ のとき，(2) と (3) より，p_1 の接尾辞は $a_1b_1wa_2b_2w'$ かつ a_1b_1w である．$w' = w_1w$ とおくと，$a_1b_1wa_2b_2w' = a_1b_1wa_2b_2w_1w$ と表すことが

できる．よって，w_1 の接尾辞は a_1b_1 となる．(1') と (2') より，p_2 の接頭辞は $wa_2b_2w'a_3$ かつ $w'a_3$ である．$w' = w_1w$ とおくと，$wa_2b_2w'a_3 = wa_2b_2w_1wa_3$ となる．さらに，$w' = ww_2$ とおくと，$w'a_3 = ww_2a_3$ と表すことができる．$|a_2b_2w_1| = |w_2a_3| + 1$ より，w_1 の後ろから 2 文字目は a_3 となる．よって，w_1 の接尾辞は a_1b_1 であり，$a_1 = a_3$ となる．これは，$a_i \neq a_j (i \neq j)$ であることに矛盾する．

$|w| > |w'|+1$ のとき，(1') と (2') より，p_2 の接頭辞は $wa_2b_2w'a_3$ かつ $w'a_3$ である．$w = w'w_1$ とおくと，$wa_2b_2w'a_3 = w'w_1a_2b_2w'a_3$ と表すことができる．このとき，w_1 の接頭辞は a_3 となる．(2) と (3) より，p_1 の接尾辞は $a_1b_1wa_2b_2w'$ かつ a_1b_1w である．$w = w'w_1$ とおくと，$a_1b_1wa_2b_2w' = a_1b_1w'w_1a_2b_2w'$ となる．さらに，$w = w_2w'$ とおくと，$a_1b_1w = a_1b_1w_2w'$ と表すことができる．$|w_1a_2b_2| = |a_1b_1w_2|$ より，w_1 の接頭辞は a_1b_1 となる．よって，w_1 の接頭辞は a_3 であり，a_1b_1 である．すなわち，$a_3 = a_1$ となる．これは，$a_i \neq a_j (i \neq j)$ であることに矛盾する．

(d-2) $q = q_1a_1b_1b_2yq_2$ $(a_2 = b_1, a_3 = b_2)$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq b_2yq_2$
(2)	$p_1 \preceq q_1a_1$	(2')	$p_2 \preceq yq_2$
(3)	$p_1 \preceq q_1a_1b_1$	(3')	$p_2 \preceq q_2$

(2) と (3) より，p_1 の接尾辞は a_1b_1 かつ a_1 である．よって，$b_1 = a_1$ となる．$a_2 = b_1$ より，$a_1 = a_2$ となり，$a_i \neq a_j (i \neq j)$ であることに矛盾する．

(d-3) $q = q_1a_1b_1b_2wa_3yq_2$ $(b_1 = a_2)$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq b_2wa_3yq_2$
(2)	$p_1 \preceq q_1a_1$	(2')	$p_2 \preceq wa_3yq_2$
(3)	$p_1 \preceq q_1a_1b_1b_2w$	(3')	$p_2 \preceq q_2$

$w = \varepsilon$ のとき，(2) と (3) より，p_1 の接尾辞は a_1 かつ $a_1b_1b_2$ である．よって，$a_1 = b_2$ となる．(1') と (2') より，p_2 の接頭辞は b_2a_3 かつ a_3 である．よって，$b_2 = a_3$ となる．したがって，$a_1 = b_2$ より，$a_1 = a_3$ となり，$a_i \neq a_j (i \neq j)$ であることに矛盾する．

$|w| \geq 1$ のとき，(2) と (3) より，p_1 の接尾辞は a_1 かつ $a_1b_1b_2w$ である．よって，w の接尾辞は a_1 となる．(1') と (2') より，p_2 の接頭辞は b_2a_3 かつ a_3 である．よって，w の接尾辞は a_3 となる．したがって，w の接尾辞は $a_1 = a_3$ となり，$a_i \neq a_j (i \neq j)$ であることに矛盾する．

(d-4) $q = q_1a_3ywa_1b_1b_2q_2$ $(b_1 = a_2)$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq wa_1b_1b_2q_2$
(2)	$p_1 \preceq q_1a_3yw$	(2')	$p_2 \preceq b_2q_2$
(3)	$p_1 \preceq q_1a_3ywa_1$	(3')	$p_2 \preceq q_2$

正規パターン p_1' と p_1'' が存在して，$p_1 = p_1'p_1''$，$p_1' \preceq q_1a_3$ かつ $p_1'' \preceq ywa_1$ が成り立つ．これらより，$p = p_1xp_2 = p_1'p_1''xp_2 \preceq q_1a_3p_1''xwa_1b_1b_2q_2 = q\{y := p_1''x\}$ となるので，$p \preceq q$ となり，$p\{x := xy\} \preceq q$ である．これは仮定に矛盾する．

(d-5) $q = q_1a_1b_1ywa_2b_2q_2$ $(b_1 = a_3)$ とする．これに対して，次の式が成り立っているものとする．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq ywa_2b_2q_2$
(2)	$p_1 \preceq q_1a_1$	(2')	$p_2 \preceq wa_2b_2q_2$
(3)	$p_1 \preceq q_1a_1b_1yw$	(3')	$p_2 \preceq q_2$

$q_1' = q_1a_1b_1$，$q_2' = yw$，$q_3' = a_2b_2q_2$ とおくと，(3) から，$p_1 \preceq q_1'q_2'$，(1') から $p_2 \preceq q_2'q_3'$ が得られ，さらに q_2' は変数記号が含まれるので，補題 6 より，$p \preceq q$ となり，$p\{x := xy\} \preceq q$ である．これは仮定に矛盾する． □

補題 12. $\sharp\Sigma \geq 3$ とし，p, q を正規パターンとする．正規パターンの有限集合 D が，次の (i), (ii) のいずれかで表されるとき，すべての $r \in D$ に対して $p\{x := r\} \preceq q$ ならば，$p\{x := xy\} \preceq q$ である．

(i) $\{a_1b_1, a_1b_2, a_2y, yb_3\}$

(ii) $\{a_1b_1, a_2b_1, yb_2, a_3y\}$
$(i \neq j$ に対して $a_i \neq a_j$かつ $b_i \neq b_j)$

証明. p に変数記号が現れない場合は自明である．したがって，$p = p_1xp_2$ (p_1, p_2 は正規パターン, x は変数記号) とおく．$p\{x := xy\} \not\preceq q$ と仮定して，矛盾を導く．

(i) $p' = p\{x := a_1 y\} = p_1 a_1 y p_2$ とおくと，$p\{x := a_1 b_1\} \preceq q$ かつ $p\{x := a_1 b_2\} \preceq q$ より，$p'\{y := b_1\} \preceq q$ かつ $p'\{y := b_2\} \preceq q$ となる．また，$p\{x := yb_3\} \preceq q$ のとき，$p\{x := a_1 b_3\} \preceq q$ であるから，$p'\{y := b_3\} \preceq q$ となる．b_i $(i = 1, 2, 3)$ は互いに異なる定数であるため，補題 7 より，$p' \preceq q$ となる．したがって，$p' = p\{x := a_1 y\} \preceq q$ かつ $p\{x := a_2 y\} \preceq q$ となり，補題 11 (i) より，$p\{x := xy\} \preceq q$ となり，仮定に矛盾する．

(ii) (i) と同様に示すことができる． □

補題 13 (Sato et al.[1])**.** $\sharp\Sigma \geq 3$ とし，p, q を正規パターンとする．このとき，ある $a \in \Sigma$ に対して，$p\{x := a\} \preceq q$ かつ $p\{x := xy\} \preceq q$ ならば $p \preceq q$ である．ただし，y は q に含まれない変数記号である．

補題 14. $k \geq 3$, $\sharp\Sigma \geq k + \sqrt{k+1}$, $P \in \mathcal{RP}^+$, $Q \in \mathcal{RP}^k$ とする．すべての定数記号 $a, b \in \Sigma$ に対し，ある正規パターン $q \in Q$ が存在し，$p\{x := ab\} \preceq q$ ならば，$p\{x := xy\} \preceq q$ となる．

証明. m を整数とし，$\sharp\Sigma = k + m$ とおく．$m \geq \sqrt{k+1}$ のとき，本補題が成り立つことを示す．

p に変数記号が現れない場合は自明である．したがって，$p = p_1 x p_2$ (p_1, p_2 は正規パターン，x は変数記号) とおく．また，$Q = \{q_1, \ldots, q_k\}$ とする．次のように記号を定める．

$$
\begin{aligned}
A_i &= \{a \in \Sigma \mid p\{x := ay\} \preceq q_i,\ y \in X\},\\
B_i &= \{b \in \Sigma \mid p\{x := yb\} \preceq q_i,\ y \in X\},\\
A &= \bigcup_{i=1}^{k} A_i,\\
B &= \bigcup_{i=1}^{k} B_i,\\
\tilde{A} &= \Sigma \setminus A,\\
\tilde{B} &= \{\bar{b} \mid b \in \Sigma \setminus B\},\\
\bar{\Sigma} &= \{\bar{c} \mid c \in \Sigma\} \quad (i = 1, \cdots, k).
\end{aligned}
$$

$p\{x := xy\} \not\preceq q_i \quad (i = 1, \cdots, k)$ と仮定する．ある i に対して，$\sharp A_i \geq 2$ かつ $\sharp B_i \geq 2$ のとき，補題 11 (i) により，$p\{x := xy\} \preceq q_i$ となる．これは仮定に矛盾する．よって，全ての i $(i = 1, \cdots, k)$ に対して，$\sharp A_i \leq 1$ かつ $\sharp B_i \leq 1$ となる．したがって，$0 \leq \sharp A \leq k$ かつ $0 \leq \sharp B \leq k$ となる．

$G = (\Sigma, \bar{\Sigma}; \tilde{A} \times \tilde{B})$ を Σ と $\bar{\Sigma}$ を部集合とし，辺集合を $\tilde{A} \times \tilde{B}$ とする 2 部グラフとする．Σ と $\bar{\Sigma}$ を部集合とし，$\tilde{E}_i = \{(a, \bar{b}) \in \tilde{A} \times \tilde{B} \mid p\{x := ab\} \preceq q_i\}$ を辺集合とする 2 分グラフを $\tilde{G}_i = (\Sigma, \bar{\Sigma}; \tilde{E}_i)$ とする．

$k \geq 3$ のとき，$m \geq \sqrt{k+1}$ であれば，$m \geq 2$ であるから，任意の $a \in A$ に対して，2 頂点 $b_a, b'_a \in \tilde{B}$ $(b_a \neq b'_a)$ が存在して，ある i $(1 \leq i \leq k)$ に対して，$p\{x := ab_a\} \preceq q_i$ かつ $p\{x := ab'_a\} \preceq q_i$ となる．また同様に，任意の $b \in B$ に対して，2 頂点 $a_b, a'_b \in \tilde{A}$ $(a_b \neq a'_b)$ が存在して，ある i $(1 \leq i \leq k)$ に対して，$p\{x := a_b b\} \preceq q_i$ かつ $p\{x := a'_b b\} \preceq q_i$ となる．$\sharp A > 0$ または $\sharp B > 0$ であるような任意の i $(1 \leq i \leq k)$ に対して，これらの 4 頂点 $b_a, b'_a \in \tilde{B}$ と $a_b, a'_b \in \tilde{A}$ を用いて，

$$
\begin{aligned}
E_i = &\{(a, \bar{b_a}), (a, \bar{b'_a}) \mid a \in A_i, b_a, b'_a \in \tilde{B}\}\\
&\cup \{(a_b, \bar{b}), (a'_b, \bar{b}) \mid b \in B_i, a_b, a'_b \in \tilde{A}\}
\end{aligned}
$$

とおく．$G_i = (\Sigma, \bar{\Sigma}; E_i \cup \tilde{E}_i)$ とする．$E_i \cap \tilde{E}_i = \emptyset$ であることに注意する．このとき，G_i の各頂点の次数は 2 以下となる．もし次数 3 以上の頂点が存在すれば，補題 7 より，$p \preceq q$ となり，$p\{x := xy\} \preceq q$ であるため，仮定に矛盾する．G_i $(1 \leq i \leq k)$ の辺の総数を見積もる．

$$
\begin{aligned}
&\sum_{i=1}^{k} \sharp(E_i \cup \tilde{E}_i)\\
&= \sharp\tilde{A} \times \sharp\tilde{B} + 2\sharp A + 2\sharp B\\
&= (k + m - \sharp A)(k + m - \sharp B) + 2\sharp A + 2\sharp B\\
&= (k+m)^2 - (\sharp A + \sharp B)(k+m)\\
&\quad + \sharp A \times \sharp B + 2\sharp A + 2\sharp B\\
&= (k+m)^2 - (\sharp A + \sharp B + 4)(k+m)\\
&\quad + (\sharp A + 2)(\sharp B + 2) - 4\\
&= (k+m)^2 - (\sharp A + \sharp B + 4)(k+m)\\
&\quad + (\sharp A + 2)(\sharp B + 2) - 4 + 4(k+m)
\end{aligned}
$$

$$= (k+m-(\sharp A+2))(k+m-(\sharp B+2))$$
$$+4(k+m)-4$$
$$= (k+m-(\sharp A+2))(k+m-(\sharp B+2))$$
$$+4(k+m)-4$$
$$\geq (m-2)^2+4(k+m)-4 = m^2+4k$$

よって，$m \geq \sqrt{k+1}$ のとき，

$$\sum_{i=1}^{k} \sharp(E_i \cup \tilde{E}_i) \geq 5k+1$$

となるので，$\sharp(E_i \cup \tilde{E}_i) \geq 6$ となる i が存在する．

$\sharp E_i = 0$，$\sharp\tilde{E}_i = 6$ のとき，G_i の各頂点の次数は 2 以下であるので，3 本の互いに隣接しない辺が存在する．補題 11(ii) 証明 (abc) より，$p\{x := xy\} \preceq q_i$ となる．これは仮定に矛盾する．

$\sharp E_i = 2$，$\sharp\tilde{E}_i = 4$ のとき，G_i の頂点の次数は 2 以下であるので，2 本の互いに隣接しない辺が存在する．補題 11(ii) 証明 (d) より，$p\{x := xy\} \preceq q_i$ となる．これは仮定に矛盾する．

$\sharp E_i = 4$，$\sharp\tilde{E}_i = 2$ のとき，G_i の頂点の次数は 2 以下であるので，2 本のパスまたは長さ 2 のパスが存在する．2 本のパスを持つ場合，1 本の互いに隣接しない辺がある．よって，補題 11(ii) 証明 (d) より，$p\{x := xy\} \preceq q_i$ となる．これは仮定に矛盾する．長さ 2 のパスを持つ場合，補題 12 より，$p\{x := xy\} \preceq q_i$ となる．これは仮定に矛盾する．

以上より，$p\{x := xy\} \preceq q$ である． □

定理 15. $k \geq 3$，$\sharp\Sigma \geq 2k-1$，$P \in \mathcal{RP}^+$，$Q \in \mathcal{RP}^k$ とする．このとき，以下の (i),(ii),(iii) は同値である．

$$\begin{array}{ll} \text{(i)} & S_2(P) \subseteq L(Q), \\ \text{(ii)} & P \sqsubseteq Q, \\ \text{(iii)} & L(P) \subseteq L(Q). \end{array}$$

証明. (ii) ⇒ (iii) と (iii) ⇒ (i) は自明である．定理 9 より，$\sharp\Sigma \geq 2k+1$ のとき，(i) ⇒ (ii) は成り立つ．よって，$\sharp Q = k$ のとき，$\sharp\Sigma = 2k-1$ または $\sharp\Sigma = 2k$ の場合，(i) ⇒ (ii) が成り立つことを，p に含まれる変数記号の数 n に関する数学的帰納法により証明する．

$n = 0$ のとき，$S_2(p) = \{p\}$ であり，$p \in L(Q)$ となる．よって，ある $q \in Q$ に対して，$p \preceq q$ となる．

$n \geq 0$ 個の変数記号を含む全ての正規パターンに対して題意が成り立つと仮定する．p を $S_2(p) \subseteq L(Q)$ を満たす $n+1$ 個の変数記号を含む正規パターンとする．$p \not\preceq q_i$ $(1 \leq i \leq k)$ と仮定する．$p = p_1 x p_2$ (p_1, p_2 は正規パターン，x は変数記号)，$Q = \{q_1, \ldots, q_k\}$ を考える．$a, b \in \Sigma$ に対して，$p_a = p\{x := a\}$ と $p_{ab} = p\{x := ab\}$ とおく．このとき，p_a, p_{ab} は n 個の変数記号が含まれ，$S_2(p_a) \subseteq L(Q)$ かつ $S_2(p_{ab}) \subseteq L(Q)$ が成り立つことに注意する．帰納法の仮定より，全ての $a, b \in \Sigma$ に対して，$p_a \preceq q_i$ かつ $p_{ab} \preceq q_{i'}$ を満たすような $i, i' \leq k$ が存在する．$D_i = \{a \in \Sigma \mid p\{x := a\} \preceq q_i\}$ $(1 \leq i \leq k)$ とする．ある i に対して，$\sharp D_i \geq 3$ であるとき，補題 7 より，$p \preceq q_i$ となる．これは仮定に矛盾する．よって，$\sharp D_i \leq 2$ $(1 \leq i \leq k)$ となる場合を考える．$\sharp\Sigma = 2k-1$ のとき，全ての i に対して，$\sharp D_i = 2$ または $\sharp D_i = 1$，$\sharp\Sigma = 2k$ のとき，全ての i に対して，$\sharp D_i = 2$ となる．補題 14 より，$p\{x := xy\} \preceq q_{i_0}$ となる i_0 が存在する．したがって，補題 13 より，$p \preceq q_{i_0}$ となる．これは仮定に矛盾する．

以上より，(i) ⇒ (ii) が成り立つ． □

系 16. $k \geq 3$，$\sharp\Sigma \geq 2k-1$，$P \in \mathcal{RP}^+$ とする．このとき，$S_2(P)$ は $\mathcal{RPL}^k$ における $L(P)$ の特徴集合である．

補題 17 (Sato et al.[1])**.** $\sharp\Sigma \leq 2k-2$ とする．このとき，$\mathcal{RP}^k$ は包含に関するコンパクト性を持たない．

次の例は，$k = 2$ における定理 15 の反例である．

例 1. $\Sigma = \{a, b, c\}$ を 3 つの定数からなる集合，p, q_1, q_2 を正規パターン，x, x', x'' を変数記号とする．

$$\begin{array}{rcl} p &=& x'axbx'', \\ q_1 &=& x'abx'', \\ q_2 &=& x'cx''. \end{array}$$

$w \in \Sigma^+$ とする．w に c が含まれるとき，$p\{x := w\} \preceq q_2$ となり，c が含まれないとき，$p\{x := w\} \preceq$

q_1 となる．よって，$L(p) \subseteq L(q_1) \cup L(q_2)$ である．しかし，$p \not\preceq q_1$ かつ $p \not\preceq q_2$ である．

定理 15 と補題 17 より，次の定理が得られる．

定理 18. $k \geq 3$ とし，$\sharp\Sigma \geq 2k-1$ とする．このとき，$\mathcal{RP}^k$ は包含に関してコンパクト性を持つ．

定理 19. $\sharp\Sigma \geq 4$ とし，$P \in \mathcal{RP}^+$，$Q \in \mathcal{RP}^2$ とする．このとき，以下の (i),(ii),(iii) は同値である．

$$\begin{array}{ll} \text{(i)} & S_2(P) \subseteq L(Q), \\ \text{(ii)} & P \sqsubseteq Q, \\ \text{(iii)} & L(P) \subseteq L(Q) \end{array}$$

証明．(ii) $\Rightarrow$ (iii) と (iii) $\Rightarrow$ (i) は自明に成り立つ．よって，(i) $\Rightarrow$ (ii) が成り立つことを示す．$Q = \{q_1, q_2\}$ とするとき，p に含まれる変数記号の数 n に関する数学的帰納法で示す．

(1) $n = 0$ のとき，p は定数記号列となるので $S_2(p) = \{p\}$ となり (i) より，$p \in L(Q)$ となる．よって，ある $q \in Q$ に対して $p \preceq q$ となる．

(2) $n = k$ 個の変数記号を含むすべての正規パターンに対して有効であると仮定する．そして，p を $S_2(p) \subseteq L(Q)$ を満たす $(n+1)$ 個の変数記号を含む正規パターンとする．

$p \not\preceq q_i$ $(i = 1, 2)$ とし，$p = p_1 x p_2$ $(p_1, p_2$ は正規パターン，x は変数記号$)$ を考える．$a, b \in \Sigma$ に対して，$p_a = p\{x := a\}$，$p_{ab} = p\{x := ab\}$ とおく．このとき，p_a, p_{ab} は n 個の変数記号が含まれ，$S_2(p_a) \subseteq L(Q)$，$S_2(p_{ab}) \subseteq L(Q)$ が成り立つことに注意する．帰納法の仮定より，全ての $a, b \in \Sigma$ に対して，$p_a \preceq q_i, p_{ab} \preceq q_{i'}$ を満たすような $i, i' \leq k$ が存在する．次のように記号を定める．

$$\begin{array}{rcl} A_i &=& \{a \in \Sigma \mid p\{x := ay\} \preceq q_i\}, \\ B_i &=& \{b \in \Sigma \mid p\{x := yb\} \preceq q_i\}, \\ \tilde{A} &=& \Sigma \backslash (A_1 \cup A_2), \\ \tilde{B} &=& \{\bar{b} \mid b \in \Sigma \setminus (B_1 \cup B_2)\}, \\ \bar{\Sigma} &=& \{\bar{a} \mid a \in \Sigma\} \quad (i = 1, 2), \\ D_i &=& \{a \in \Sigma \mid p\{x := a\} \preceq q_i\} \quad (i = 1, 2). \end{array}$$

ある i に対して $\sharp D_i \geq 3$ のとき，補題 7 より，$p \preceq q_i$ となる．よって，全ての i に対して，$\sharp D_i \leq 2$ となる．したがって，$\sharp D_1 = 2$ かつ $\sharp D_2 = 2$ となる場合を考える．

ある i に対して，$\sharp A_i \geq 2$ または $\sharp B_i \geq 2$ のとき，補題 11(i)，補題 13 より，$p \preceq q$ となる．これは仮定に矛盾する．よって，全ての i に対して，$\sharp A_i \leq 1$ かつ $\sharp B_i \leq 1$ となる．$E_i = \{(a, \bar{b}) \in \tilde{A} \times \tilde{B} \mid p\{x := ab\} \preceq q_i\}$ とする 2 分グラフ $G_i = (\tilde{A}, \tilde{B}; E_i)$ を定義する．このとき，G_i の各頂点の次数は，2 以下となる．3 以上の場合，補題 7 より，$p \preceq q_i$ となり，仮定に矛盾する．

$\sharp A_i = 0$ かつ $\sharp B_i = 0$ $(i = 1, 2)$ のとき，$\sharp E_1 + \sharp E_2 = 4 \times 4 = 16$ となる．このとき，$\sharp E_{i_0} \geq 8$ となる i_0 が存在する．$\sharp E_{i_0} = 8$ のとき，G_{i_0} は 2 つの長さ 4 のサイクルを含む．よって，4 本の互いに隣接しない辺が存在する．補題 11(ii)，補題 13 より，$p \preceq q_{i_0}$ となる．これは仮定に矛盾する．

$\sharp A_1 = 1$，$\sharp A_2 = 0$ かつ $\sharp B_i = 0$ $(i = 1, 2)$ のとき，$\sharp E_1 + \sharp E_2 = 3 \times 4 = 12$ となる．このとき，$\sharp E_{i_0} \geq 6$ となる i_0 が存在する．$\sharp E_{i_0} = 6$ のとき，G_{i_0} は長さ 4 のサイクルと長さ 3 のパスまたは長さ 6 のパスを含む．よって，3 本の互いに隣接しない辺が存在する．したがって，補題 11(ii) 証明 (abc)，補題 13 より，$p \preceq q_{i_0}$ となる．これは仮定に矛盾する．

$\sharp A_1 = 1$，$\sharp A_2 = 1$ かつ $\sharp B_i = 0$ $(i = 1, 2)$ のとき，$\sharp E_1 + \sharp E_2 = 2 \times 4 = 8$ となる．このとき，$\sharp E_{i_0} \geq 4$ となる i_0 が存在する．$\sharp E_{i_0} = 6$ のとき，G_{i_0} は長さ 4 のサイクルまたは長さ 4 のパスを含む．よって，2 本の互いに隣接しない辺が存在する．したがって，補題 11(ii) 証明 (d)，補題 13 より，$p \preceq q_{i_0}$ となる．これは仮定に矛盾する．

$\sharp A_1 = 1$，$\sharp A_2 = 0$，$\sharp B_1 = 1$ かつ $\sharp B_2 = 0$ のとき，$\sharp E_1 + \sharp E_2 = 3 \times 3 = 9$ となる．このとき，$\sharp E_{i_0} \geq 5$ となる i_0 が存在する．$\sharp E_{i_0} = 5$ のとき，G_{i_0} は長さ 5 のパスまたは長さ 4 のサークルと長さ 1 のパスを含む．よって，3 本の互いに隣接しない辺が存在する．したがって，補題 11(ii) 証明 (abc)，補題 13 より，$p \preceq q_{i_0}$ となる．これは仮定に矛盾する．

$\sharp A_1 = 1$，$\sharp A_2 = 1$，$\sharp B_1 = 1$ かつ $\sharp B_2 = 0$ のとき，$\sharp E_1 + \sharp E_2 = 2 \times 3 = 6$ となる．このとき，$\sharp E_{i_0} \geq 3$ となる i_0 が存在する．$\sharp E_{i_0} = 3$ のとき，G_{i_0} は長さ 3 のパスまたは長さ 1 のパスと長さ 2 のパスを含む．よって，2 本の互いに隣接しない辺が存在する．したがって，補題 11(ii) 証明 (d)，補題 13 より，$p \preceq q_{i_0}$ となる．これは仮定に矛盾する．

$\sharp A_i = 1$ かつ $\sharp B_i = 1$ $(i = 1, 2)$ のとき，$\sharp E_1 + \sharp E_2 = 2 \times 2 = 4$ となる．このとき，$\sharp E_{i_0} \geq 2$ となる i_0 が存在する．$\sharp E_{i_0} = 3$ のとき，G_{i_0} は長さ 2 のパスまたは 2 本の長さ 1 のパスを含む．長さ 2 のパスの場合，ab_j または $a_j b$ $(j = 1, 2)$ となる文字列となる．補題 12 より，$p\{x := xy\} \preceq q_{i_0}$ となる．これは仮定に矛盾する．2 本の長さ 1 のパスの場合，2 本の互いに隣接しない辺が存在する．補題 11(ii) 証明 (d)，補題 13 より，$p \preceq q_{i_0}$ となる．これは仮定に矛盾する．

以上より，(i) $\Rightarrow$ (ii) が成り立つ． □

定理 19 より，次の 2 つの系が成り立つ．

系 20. $\sharp\Sigma \geq 4$ とし，$P \in \mathcal{RP}^+$ とする．このとき，$S_2(P)$ は，$\mathcal{RPL}^2$ における $L(P)$ の特徴集合である．

系 21. $\sharp\Sigma \geq 4$ とする．このとき，クラス $\mathcal{RP}^2$ は包含に関するコンパクト性を持つ．

4 非隣接変数正規パターン

隣接した変数記号を持たない正規パターンを**非隣接変数正規パターン**という．$\mathcal{RP}_{\mathcal{NAV}}$ を非隣接変数正規パターン全体の集合とする．このとき，次の定理が成り立つ．

定理 22. $\sharp\Sigma \geq k + \sqrt{k+1}$ とし，$P \in \mathcal{RP}^+_{\mathcal{NAV}}$，$Q \in \mathcal{RP}^k_{\mathcal{NAV}}$ とする．このとき，以下の (i), (ii), (iii) は同値である．

(i) $S_2(P) \subseteq L(Q)$,
(ii) $P \sqsubseteq Q$,
(iii) $L(P) \subseteq L(Q)$

証明．定義より，(ii) $\Rightarrow$ (iii) と (iii) $\Rightarrow$ (i) は自明に成り立つ．よって，(i) $\Rightarrow$ (ii) が成り立つことを，p に現れる変数記号の数 n に関する数学的帰納法で証明する．$n = 0$ のとき，$S_2(p) = \{p\}$ であり，$p \in L(Q)$ となる．よって，ある $q \in Q$ に対して，$p \preceq q$ となる．$n \geq 0$ 個の変数記号を含む全ての正規パターンに対して，題意が成り立つと仮定する．p を $S_2(p) \subseteq L(Q)$ を満たす $n+1$ 個の変数記号を含む非隣接変数正規パターンとする．$p \not\preceq q_i$ $(i = 1, 2)$ と仮定する．非隣接変数正規パターン p を $p = p_1 x p_2$, $Q = \{q_1, \ldots, q_k\}$ とおく．ここで，p_1 は末尾が定数記号である非隣接変数正規パターンであり，p_2 は先頭が定数記号である非隣接変数正規パターン，x は変数記号，各 i $(1 \leq i \leq k)$ に対して，q_i は非隣接変数正規パターンである．$a, b \in \Sigma$ に対して，$p_a = p\{x := a\}$，$p_{ab} = p\{x := ab\}$ とおく．このとき，p_a, p_{ab} は n 個の変数記号が含まれ，$S_2(p_a) \subseteq L(Q)$ かつ $S_2(p_{ab}) \subseteq L(Q)$ が成り立つことに注意する．帰納法の仮定より，全ての $a, b \in \Sigma$ に対して，$p_a \preceq q_i$ かつ $p_{ab} \preceq q_{i'}$ を満たすような $i, i' \leq k$ が存在する．

補題 14 より，ある i に対して $p\{x := xy\} \preceq q_i$ が成り立つ．このとき，$p\{x := xy\} = p_1 xy p_2$ の部分パターン xy は q_i の変数記号を置き換えることで生成できない．このことは，q_i に xy が含まれることを示している．これは，q_i が非隣接変数正規パターンであることに矛盾する．

以上より，(i) $\Rightarrow$ (ii) が成り立つ． □

5 おわりに

本稿では、高々$k (k \geq 3)$ 個の正規パターン集合全体のクラス $\mathcal{RP}^k$ について、(1) 正規パターン集合 $P \in \mathcal{RP}^k$ から得られる記号列の集合 $S_2(P)$ が P により生成される言語 $L(P)$ の特徴集合となること、および (2) $\mathcal{RP}^k$ が包含に関してコンパクト性を持つことを示した Sato ら [1] の結果の証明の誤りを修正した。次に、隣接する変数がない正規パターンである非隣接変数正規パターンについて、高々$k (k \geq 3)$ 個の非隣接変数正規パターン集合全体のクラス $\mathcal{RP}^k_{\mathcal{NAV}}$

から得られる記号列の集合 $S_2(P)$ が、正規パターン言語の有限和に対する特徴集合と、定数記号の数が $k+\sqrt{k+1}$ 以上のとき、$\mathcal{RP}^k_{\mathcal{NAV}}$ が包含に関してコンパクト性をもつことを示した。これらにより、Arimura ら [5] が示した $\mathcal{RP}^k$ に対する学習アルゴリズムを非隣接変数正規パターン言語の有限和に関する効率的な学習アルゴリズムが設計できることを示した。

今後の課題として、$\mathcal{RP}^k_{\mathcal{NAV}}$ に対する特徴集合を活用し、非隣接変数正規パターン言語の有限和を正例から極限同定する多項式時間帰納推論アルゴリズムおよび一つの正例と多項式回の所属性質問を用いて同定する質問学習アルゴリズムの高速化が考えられる。また、正規パターン言語の有限和に対する特徴集合の概念を線形項木パターン言語 [6] の有限和や正則 FGS 言語 [7] に拡張することが考えられる。

謝辞

本研究は JSPS 科研費 19K12103, 21K12021, 22K12172 の助成を受けたものである。

参考文献

[1] M.Sato, Y.Mukouchi, D.Zheng, Characteristic Sets for Unions of Regular Pattern Languages and Compactness, in Proc. ALT '98, LNAI 1501, pp.220-233, 1998.

[2] Y. Mukouchi, Characterization of Pattern Languages, in Workshop ALT '91, Ohmusha, pp.93-104, 1991.

[3] Y.Mukouchi, Characterization of Pattern Languages, in Proc.ALT '98, LNAI 1501, pp.93-104, 1991.

[4] K.Wright, Identification of Unions of Languages Drawn from an Identifiable Class, in Proc.COLT '89, pp.328-333, 1989.

[5] H. Arimura, T. Shinohara, S. Otsuki, Finding Minimal Generalizations for Unions of Pattern Languages and Its Application to Inductive Inference from Positive Data, in Proc. STACS '94, LNCS 775, pp.649-660, 1994.

[6] Y. Suzuki, Ordered Term Tree Languages Which are Polynomial Time Inductively Inferable from Positive Data, Theoretical Computer Science, 350, pp.63-90, 2006.

[7] T. Uchida, T. Shoudai, S. Miyano, Parallel Algorithms for Refutation Tree Problem on Formal Graph Systems, E78-D(2), pp.99-112, 1995.