PAPER

Compactness of Finite Union of Regular Patterns and Regular Patterns without Adjacent Variables

Naoto TAKETA†, *Nonmember*, Tomoyuki UCHIDA†, Takayoshi SHOUDAI††, Satoshi MATSUMOTO†††, Yusuke SUZUKI†, *and* Tetsuhiro MIYAHARA†, *Members*

SUMMARY 正規パターンとは，定数記号と変数記号から成る，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 個の正規パターンの集合全体のクラス $\mathcal{RP}^k$ に属する正規パターン集合 P と Q に対し，任意の正規パターン $p \in P$ に対して，p より汎化された正規パターン q が Q に存在するとき，$P \sqsubseteq Q$ と書く．高々k $(k \geq 2)$ 個の正規パターンから成る集合の全体のクラスを $\mathcal{RP}^k$ 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．次に，定数記号の数が $2k-1$ 以上のとき，$\mathcal{RP}^k$ が包含関係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 $\mathcal{RP}_{NAV}$ を与え，高々k $(k \geq 1)$ 個の非隣接変数正規パターンから成る集合の全体のクラス $\mathcal{RP}^k_{NAV}$ 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 $k+2$ 以上のとき，$\mathcal{RP}^k_{NAV}$ 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数の定数記号で，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．

key words: *Regular Pattern Language, Compactness*

1. Introduction

パターンとは，定数記号と変数記号から成る記号列である．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．パターンから成る集合の全体を $\mathcal{P}$ で表す．パターン $p \in \mathcal{P}$ 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 $\}$ を表す．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．正規パターンから成る全体の集合を $\mathcal{RP}$ で表す．パターン $p \in \mathcal{P}$ がパターン $q \in \mathcal{P}$ 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．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 $p, q \in \mathcal{P}$ に対して，$p \preceq q$ ならば $L(p) \subseteq L(q)$ であ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，定数記号の数が 3 以上の場合，任意の正規パターン $p, q \in \mathcal{RP}$ に対して，$L(p) \subseteq L(q)$ ならば $p \preceq q$ も成り立つことを示した．

$\mathcal{RP}^+$ を $\mathcal{RP}$ の空でない有限集合の集合とする．$\mathcal{RP}^k$ を高々k $(k \geq 2)$ 個の正規パターンから成る集合の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 $\{L(P) \mid P \in \mathcal{RP}^k\}$ を $\mathcal{RPL}^k$ とする．$P, Q \in \mathcal{RP}^k$ とし，$Q = \{q_1, \ldots, q_k\}$ とする．任意の正規パターン $p \in P$ に対し，ある正規パターン q_i が存在し，$p \preceq q_i$ が成り立つとき $P \sqsubseteq Q$ と書く．定義より，$P \sqsubseteq Q$ 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，$k \geq 3$ 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 の定数記号列を代入することで $P \in \mathcal{RP}^k$ から得られる定数記号列の有限集合 $S_2(P)$ 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 $L(P) \subseteq L'$ となることを示し，$(i) S_2(P) \subseteq L(Q)$, (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持つことも示した．これに対し，本稿では，隣接した変数記号（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 $\mathcal{RP}_{NAV}$ 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 $\mathcal{RP}^k_{NAV}$ 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 $k+2$ 以上のとき，$\mathcal{RP}^k_{NAV}$ が包含に関してコンパクト性を持つことを示した．表 1 に本稿の結果をまとめて示す．

Table 1 包含に関してコンパクト性を持つための定数記号の数に関する条件

k	2	3 以上
$\mathcal{RP}^k$	4 以上	$2k-1$ 以上
$\mathcal{RP}^k_{NAV}$	$k+2$ 以上	

本稿の結果は，言語の有限和の表現である正規パ

Manuscript received January 1, 2015.
Manuscript revised January 1, 2015.
†Graduate School of Information Sciences, Hiroshima City University
††Department of Computer Science and Engineering, Fukuoka Institute of Technology
†††Faculty of Science, Tokai University
DOI: 10.1587/trans.E0.??.1

ターンの集合あるいは非隣接変数正規パターンの集合を対象とした効率的な学習アルゴリズムをそれぞれ与えられることを示唆している．

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，$\mathcal{RP}^k_{NAV}$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$\mathcal{RP}^k_{NAV}$が包含に関してコンパクト性をもつことを示す．

2. Preliminaries

Let Σ be a nonempty finite set of constant symbols and X an infinite set of variable symbols such that $\Sigma \cap X = \emptyset$. A pattern is a finite string which consists of symbols in $\Sigma \cup X$. Note that the empty string ε is a regular pattern in this paper. A pattern p is said to be *regular* if each variable symbol appears at most one in p. It is clear that ε is a regular pattern. The length of p, denote by $|p|$, is the number of symbols in p. The set of all patterns and regular patterns are denoted by $\mathcal{P}$ and $\mathcal{RP}$ respectively. For a set S, we denote by $\sharp S$ the number of elements in S. In this paper, we assume $\sharp\Sigma \geq 2$. A substitution θ is a mapping from $(\Sigma \cup X)^*$ to $(\Sigma \cup X)^*$ such that (1) θ is a homomorphism with respect to string concatenation, that is, for patterns π_1 and π_2, $\theta(\pi_1 \cdot \pi_2) = \theta(\pi_1) \cdot \theta(\pi_2)$ holds, (2) for each constant symbol $a \in \Sigma$, $\theta(a) = a$ holds, and (3) for each variable symbol $x \in X$, $|\theta(x)| \geq 1$ holds. Let $x_1, \ldots, x_n$ are variable symbols and $p_1, \ldots, p_n$ nonempty patterns. We denote by $\{x_1 := p_1, \ldots, x_n := p_n\}$ a substitution that replaces each variable symbol x_i with a nonempty pattern p_i for $i \in \{1, \ldots, n\}$ respectively. We denote by $\theta(\pi)$ a new pattern obtained by θ for a pattern π. For a pattern p and q, q is a *generalization* of p, or p is an *instance* of q, denoted by $p \preceq p$, if there exists a substitution θ sudh that $p = q\theta$. We say that p is equal to q, denoted by $p \equiv q$, if $p \preceq q$ and $p \succeq p$. For a pattern p, the *pattern language* of p, denoted by $L(p)$, is the set $\{w \in \Sigma^* \mid w \preceq p\}$. In particular, if p is a regular pattern, we say that $L(p)$ is a *regular pattern language*. The set of all pattern languages and regular patterns languages are denoted by $\mathcal{PL}$ and $\mathcal{RPL}$ respectively. For patterns p and q, it is clear that $L(p) = L(q)$ if $p \equiv q$, and $L(p) \subseteq L(q)$ if $p \preceq q$. Note that $L(\varepsilon) = \{\varepsilon\}$.

Lemma 1 (Mukouchi[2]): Let p and q be regular patterns. Then $p \preceq q$ if and only if $L(p) \subseteq L(q)$.

Next, we consider unions of pattern languages. The class of all nonempty finite subsets of $\mathcal{P}$ is denoted by $\mathcal{P}^+$, that is, $\mathcal{P}^+ = \{P \subseteq \mathcal{P} \mid 0 < \sharp P < \infty\}$. For a positive integer k, the class of nonempty sets consisting of at most k patterns, that is, $\mathcal{P}^k = \{P \subseteq \mathcal{P} \mid 0 < \sharp P \leq k\}$. We denote by $\mathcal{PL}^k$ the class of unions of at most k pattern languages, that is, $\mathcal{PL}^k = \{L(P) \mid P \in \mathcal{P}^k\}$, where $L(P) = \cup_{p \in P} L(p)$. In a similar way, we also define $\mathcal{RP}^+$, $\mathcal{RP}^k$ and $\mathcal{RPL}^k$ respectively. For P, Q in $\mathcal{P}^+$, we denote by $P \sqsubseteq Q$ if for any $p \in \mathcal{P}$, there is a pattern $q \in Q$ such that $p \preceq q$. It is clear that $P \sqsubseteq Q$ implies $L(P) \subseteq L(Q)$. However, note that the converse is not valid in general.

Definition 1: Let L be a class of languages, L a language in L and S a nonempty finite set of L. We say that S is a *characteristic* set for L within L if for any $L' \in$ L, $S \subseteq L'$ implies $L \subseteq L'$.

Let n be a positive integer and p a regular pattern. We denote by $S_n(p)$ the set of all strings in Σ^* obtained by replacing all variable symbols in p with strings in Σ^+ of which the length is at most n. Moreover, for a set $P \in \mathcal{RP}^+$, we define $S_n(P)$ as follows:

$$S_n(P) = \bigcup_{p \in P} S_n(p).$$

It is clear that $S_n(P) \subseteq S_{n+1}(P) \subseteq L(P)$ for any positive integer n.

Theorem 1 (Sato et al.[1]): Let k be a positive integer and $P \in \mathcal{RP}^k$. Then, there exists a positive integer n such that $S_n(P)$ is a characteristic set for $L(P)$ within $\mathcal{RPL}^k$.

Let p_1, p_2, r, q be regular patterns with $p_1 r p_2 \preceq q$ and $x_1, \ldots, x_n$ variable symbols appearing in q. In [2], the regular pattern r in $p_1 r p_2$ is said to be generated from q by variable substitution if there exist a variable symbol x_i and a substitution $\theta = \{x_1 := r_1, \ldots, x_i := r' r r'', \ldots, x_n := r_n\}$ such that $p_1 = (q_1\theta)r'$, $p_2 = r''(q_2\theta)$ for $q = q_1 x_i q_2$. It is clear that $p_1 x p_2 \preceq q$ if the regular pattern r in $p_1 r p_2$ is generated from q by variable substitution.

Theorem 2 (Sato et al.[1]): Let p, q, p_1, p_2, q_1, q_2, q_3 be regular patterns and x a variable symbol with $p = p_1 x p_2$ and $q = q_1 q_2 q_3$. Then $p \preceq q$ if the following three conditions are holds:

(i) $p_1 \preceq q_1 q_2$, (ii) $p_2 \preceq q_2 q_3$,
(iii) q_2 contains at least one variable symbol.

Let p_1, p_2, q be regular patterns and a a constant symbol with $p_1 a p_2 \preceq q$. If $p_1 x p_2 \npreceq q$, then the constant symbol a in $p_1 a p_2$ is not generated from q by variable substitution. Thus $q = q_1 a q_2$ for some regular patterns such that $p_1 \preceq q_1$ and $p_2 \preceq q_2$. From the above, the following lemma holds.

Lemma 2 (Sato et al.[1]): Suppose $\sharp\Sigma \geq 3$. Let p, p_1, p_2, q are regular patterns and x a variable symbol with $p = p_1 x p_2 \preceq p$ Let a, b and c be mutually distinct constant symbols. If $p_1 a p_2 \preceq q$, $p_1 b p_2 \preceq q$ and $p_1 c p_2 \preceq q$, then $p \preceq q$ holds.

Lemma 3 (Sato et al.[1]): Suppose $\sharp\Sigma \geq 3$. Let p_1, p_2, q_1, q_2 be regular patterns and x a variable symbol. Let a, b be constant symbols with $a \neq b$ and w a string in Σ^*. Let $p = p_1 AwxwB p_2$ and $q = q_1 AwB q_2$ be regular patterns satisfies the following three conditions:

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

If $p\{x := a\} \preceq q$ and $p\{x := b\} \preceq q$, then we have $p \npreceq q$.

From Lemma **??**, the following lemma holds.

Theorem 3 (Sato et al.[1]): Let $\sharp\Sigma \geq 2k+1$, $P \in \mathcal{RP}^+$ and $Q \in \mathcal{RP}^k$. Then, the following (i), (ii) and (iii) are equivalent:

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

Example 1 in [1] is given as a counter-example of Theorem 3.

From Theorem 3, we have the following corollary.

Corollary 1 (Sato et al.[1]): Let $\sharp\Sigma \geq 3$ and p, q regular patterns. Then, the following (i), (ii) and (iii) are equivalent:

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

3. 正規パターン集合のコンパクト性

この節では，コンパクト性の定義を与え，$\sharp\Sigma \geq 2k-1$ 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，$\mathcal{RP}^k$ が包含に関してコンパクト性を持つことを示す．

Definition 2: クラス $C \subseteq \mathcal{RP}^+$ が**包含に関してコンパクト性を持つ**とは，任意のパターン $p \in \mathcal{RP}$ と任意のパターン集合 $Q \in C$ 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 $C \in \mathcal{P}^+$ が包含に関してコンパクト性を持つことが定義できる．また，クラス $C \in \mathcal{RP}^+$ が包含に関してコンパクト性を持つとき，補題**??**より，任意の $P, Q \in C$ に対して，$P \sqsubseteq Q$ ならばその時に限り $L(P) \subseteq L(Q)$ であることが示せる．

Lemma 4 (Sato et al.[1]): $\sharp\Sigma \geq 3$ 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，任意の $r \in D$ に対して $p\{x := r\} \preceq q$ ならば，$p\{x := xy\} \preceq q$ である．ただし，$a \neq b$ とする．

(i) $\{ay, by\}$, (ii) $\{ya, yb\}$.

Proof. 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．$p\{x := xy\} \npreceq q$ と仮定して，矛盾を導く．

(i) $D = \{ay, by\}$ $(a \neq b)$ であるとする．$p\{x := xy\} \npreceq q$ のとき，$p_1ayp_2 \preceq q$ 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') がすべて成り立つ．

(1)	$p_1 \preceq q_1$	(1')	$p_2 \preceq wby_2q_2$ または $p_2 \preceq y'wby_2q_2$
(2)	$p_1 \preceq q_1ay_1w$	(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，$p_1 = p'_1p''_1$，$p'_1 \preceq q_1a$ かつ $p''_1 \preceq y_1w$ が成り立つ．したがって，$p = p_1xp_2 = p'_1p''_1xp_2$ であるから，(1') が $p_2 \preceq wby_2q_2$ のとき，$p \preceq q_1ap''_1xwby_2q_2 = q\{y_1 := p''_1x\}$ となる．また，(1') が $p_2 \preceq y'wby_2q_2$ のとき，$p \preceq q_1ap''_1xy'wby_2q_2 = q\{y_1 := p''_1xy'\}$ となる．よって，$p \preceq q$ が成り立ち，仮定 $p\{x := xy\} \npreceq q$ に矛盾する．

(ii) $D = \{ya, yb\}$ $(a \neq b)$ のときは，記号列 p と q を逆順にすることにより，**(i)** の場合と同様に，仮定 $p\{x := xy\} \npreceq q$ に矛盾することを証明できる．

Lemma 5: $\sharp\Sigma \geq 4$ 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 $D = \{a_1b_1, a_2b_2, a_3b_3, a_4b_4\}$ ($i \neq j$ に対して，$a_i \neq a_j$ かつ $b_i \neq b_j$) で表されるとき，任意の $r \in D$ に対して $p\{x := r\} \preceq q$ ならば，$p\{x := xy\} \preceq q$ である．

Proof. 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．$p\{x := xy\} \npreceq q$ と仮定して，矛盾を導く．

$D = \{a_1b_1, a_2b_2, a_3b_3, a_4b_4\}$ ($i \neq j$ に対して，$a_i \neq a_j$ かつ $b_i \neq b_j$) であるとする．任意の $r \in D$ 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) $a_1b_1, y_1b_2, a_3y_2, a_4y_3$
(b) $a_1b_1, a_2b_2, a_3b_3, a_4y_1$	(j) $a_1b_1, a_2y_1, a_3y_2, a_4y_3$
(c) $a_1b_1, a_2b_2, a_3b_3, y_1b_4$	(k) $y_1b_1, y_2b_2, y_3b_3, y_4b_4$
(d) $a_1b_1, a_2b_2, a_3y_1, y_2b_4$	(l) $y_1b_1, y_2b_2, y_3b_3, a_4y_4$
(e) $a_1b_1, a_2b_2, y_1b_3, y_2b_4$	(m) $y_1b_1, y_2b_2, a_3y_3, a_4y_4$
(f) $a_1b_1, a_2b_2, a_3y_1, a_4y_2$	(n) $y_1b_1, a_2y_2, a_3y_3, a_4y_4$
(g) $a_1b_1, y_1b_2, y_2b_3, y_3b_4$	(o) $a_1y_1, a_2y_2, a_3y_3, a_4y_4$
(h) $a_1b_1, y_1b_2, y_2b_3, a_4y_3$	(y_1, y_2, y_3, y_4 は変数記号)

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 4(i) または (ii) に対応する記号列が現れる．その場合の証明より仮定 $p\{x := xy\} \npreceq q$ 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 $a_1b_1, a_2b_2, a_3b_3, a_4b_4$ が現れる場合，(d) は，q に $a_1b_1, a_2b_2, a_3y_1, y_2b_4$ が現れる場合において，矛盾を導く証明が考えられる．しかし，(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_1 が現れる場合と q に a_1b_1, a_2b_2, y_2b_4 が現れる場合において，矛盾を導くことで，証明できる．よって，本論文では，q に a_1b_1, a_2b_2, a_3b_3 が現れる場合と q に a_1b_1, a_2b_2, a_3y $(y = y_1)$ が現れる場合を証明する．q に a_1b_1, a_2b_2, y_2b_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．

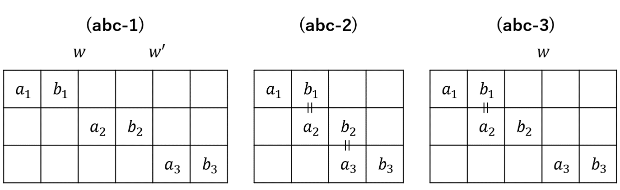

Fig. 1 (abc) の場合分け

(abc) q に a_1b_1, a_2b_2, a_3b_3 が現れる場合，

図 1 のように，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) $q = q_1a_1b_1a_3b_3q_2$ ($b_1 = a_2$ かつ $a_3 = b_2$),
(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') $p_2 \preceq q_2$

$|w| = |w'|$ 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 であるので，$a_1b_1w = a_2b_2w'$ となる．よって，$a_1b_1 = a_2b_2$ となり，$a_1 \neq a_2$ かつ $b_1 \neq b_2$ 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 $wa_2b_2w'a_3b_3$ 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，$a_3 \neq a_1$ 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′) 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．$w' = ww_2$ とおくと，$w'a_3b_3 = ww_2a_3b_3$ となり，$w' = w_1w$ とおくと，$wa_2b_2w'a_3b_3 = wa_2b_2w_1wa_3b_3$ となる．$|ww_2a_3b_3| = |wa_2b_2w_1|$ 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，$a_2 \neq a_3$ かつ $b_2 \neq b_3$ であることに矛盾する．

(abc-2) $q = q_1a_1b_1a_3b_3q_2$ ($b_1 = a_2$ かつ $a_3 = b_2$) 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，$a_1 \neq a_2$ 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 $b_2a_3b_3$

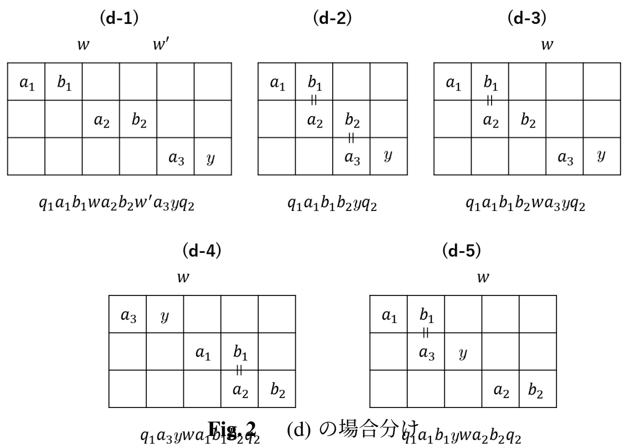

Fig. 2 (d) の場合分け

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，$a_1 \neq a_3$ 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 の最後の記号は 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 の最後の記号は a_3 となる．したがって，w の最後の記号は $a_1 = a_3$ となり，$a_1 \neq a_3$ 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，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') $p_2 \preceq q_2$

$|w| = |w'|$ のとき，(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，$p = p_1xp_2 = p'_1p''_1xp_2 \preceq q_1a_3p''_1xwa_1b_1w'a_2b_2q_2 = q\{y := p''_1x\}$ 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**??**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，表 2 のように，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) $q = q_1a_1b_1b_2yq_2$ ($a_2 = b_1$ かつ $a_3 = b_2$),
(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，$a_1 \neq a_3$ 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 の最後から 2 つ目の記号は 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，$a_1 \neq a_3$ 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．$|w_1a_2b_2w'| = |a_1b_1w_2w'|$ 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，$a_1 \neq a_3$ 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 の最初の記号は 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，$a_3 \neq a_1$ であることに矛盾する．

(d-2) $q = q_1a_1b_1b_2yq_2$ ($a_2 = b_1$ かつ $a_3 = b_2$) 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，$a_1 \neq a_2$ 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，$a_1 \neq a_3$ 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 の最後の記号は a_1 となる．(1') と (2') より，p_2 の接頭辞は b_2wa_3 かつ wa_3 である．よって，w の最後の記号は a_3 となる．したがって，w の最後の記号は $a_1 = a_3$ となり，$a_1 \neq a_3$ 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') $p_2 \preceq q_2$

(3) より，正規パターン p'_1 と p''_1 が存在して，$p_1 = p'_1p''_1$，$p'_1 \preceq q_1a_3$ かつ $p''_1 \preceq ywa_1$ が成り立つ．これらより，$p = p_1xp_2 = p'_1p''_1xp_2 \preceq q_1a_3p''_1xwa_1b_1b_2q_2 = q\{y := p''_1x\}$ 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，$p_1 \preceq q'_1q'_2$，(1') から $p_2 \preceq q'_2q'_3$ が得られ，さらに q'_2 は変数記号が含まれるので，補題**??**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．

Takayoshi Shoudai, 2024-05-13.

Lemma 6: Let Σ be an alphabet with $\Sigma \geq 3$ and p, q regular patterns on Σ. Let D be the following set of regular patterns on Σ. Then, if $p\{x := r\} \preceq q$ for all $r \in D$, then $p\{x := xy\} \preceq q$:

$$D = \{ya, bc, dy\}\ (b \notin \{a, d\} \text{ and } c \notin \{a, d\}).$$

Proof. It is obvious if no variable symbol appears in p. Therefore, let $p = p_1xp_2$, where p_1, p_2 are regular patterns and x is a variable symbol. We assume that $p\{x := xy\} \npreceq q$ in order to derive the contradiction.

Since $p\{x := xy\} \npreceq q$, there exist three regular patterns A, B, C on Σ such that $q = q_1AwBw'Cq_2$ hold, where $\{A, B, C\} = \{y_1a, bc, dy_2\}$ for some variable symbols y_1, y_2 $(y_1 \neq y_2)$. For these p_1, p_2, q_1, q_2, the following six equations hold:

(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') $p_2 \preceq q_2$

Let $q'_1 = q_1A$, $q'_2 = wBw'$, $q'_3 = Cq_2$. From (3) and (1'), $p_1 \preceq q'_1q'_2$ and $p_2 \preceq q'_2q'_3$ hold. If a variable symbol appears in q'_2, from Lemma 2, $p \preceq q$ holds. This implies that the case of either $B = y_1a$ or $B = dy_2$ contradicts the assumption. Then, B must be bc. If $A = dy_2$, (2) becomes $p_1 \preceq q_1dy_2w$. For some p'_1 and p''_1, let $p_1 = p'_1p''_1$, where $p'_1 \preceq q_1d$ and $p''_1 \preceq y_2w$. From (1'), we have $p = p_1xp_2 = p'_1p''_1xp_2 \preceq q_1dp''_1xwbcw'y_1aq_2 = q\{y_2 := p''_1x\}$. Thus, $p\{x := xy\} \preceq q\{y_2 := p''_1xy\}$ holds. This contradicts the assumption.

Below, we consider the case of $A = y_1a, B = bc$, and $C = dy_2$. Let $q = q_1y_1awbcw'dy_2q_2$, where $b \notin \{a, d\}$ and $c \notin \{a, d\}$. If $p\{x := xy\} \npreceq q$ holds, we have the following equations:

(1) $p_1 \preceq q_1$ (1') $p_2 \preceq wbcw'dy_2q_2$
(2) $p_1 \preceq q_1y_1aw$ (2') $p_2 \preceq w'dy_2q_2$
(3) $p_1 \preceq q_1y_1awbcw'$ (3') $p_2 \preceq q_2$

Note that from (1') and (2'), $wbcw'd$ and $w'd$ are prefixes of p_2, and from (2) and (3), $awbcw'$ and aw are suffixes of p_1.

If $|w| = |w'|$, then $c = a$ holds. It contradicts the condition $c \neq a$.

If $|w| = |w'| + 1$, then $b = a$ holds. It contradicts the condition $b \neq a$.

If $|w| = |w'| + 2$, since $awbcw'$ and aw are suffixes of p_1, and since $|w| \geq 2$, a is a suffix of w. From (1') and (2'), since $wbcw'd$ and $w'd$ are prefixes of p_2, we have $w = w'da$. Since $awbcw'$ and aw are suffixes of p_1, we have $w = bcw'$. Therefore, $w'da = bcw'$ holds. We show the next claim:

Claim 1. Let w' be a string of constant symbols in Σ and a, b, c, d constant symbols in Σ. Then, if $b \notin \{a, d\}$ and $c \notin \{a, d\}$, then $w'da \neq bcw'$ holds.

Proof of Claim 1. When $|w'| = 0, 1, 2, 3$, it is easy to see that $w'da = bcw'$ does not satisfy the conditions $b \notin \{a, d\}$ and $c \notin \{a, d\}$. Therefore, $w'da \neq bcw'$ holds. Let $n = |w'|$. When $n \geq 4$, we assume that for any string w'' with $|w''| < n$, if $b \notin \{a, d\}$ and $c \notin \{a, d\}$, $w''da \neq bcw''$ holds. Since the string w' has a prefix bc and a suffix da, there exists a string w'' with $|w''| \geq 0$ such that $w' = bcw''da$ holds. Since $w'da = bcw''dada$ and $bcw' = bcbcw''da$, if $w'da = bcw'$ holds, we have $bcw''dada = bcbcw''da$. Then we conclude that $w''da = bcw''$. It contradicts the induction hypothesis. Thus, $w'da \neq bcw'$ holds. From the above, for any string w' with $|w'| \geq 0$, if $b \notin \{a, d\}$ and $c \notin \{a, d\}$, $w'da \neq bcw'$ holds. (*End of Proof of Claim*)

Thus, the case of $|w| = |w'| + 2$ contradicts *Claim* 1.

If $|w| \geq |w'| + 3$, from (2) and (3), there exists a string w'' of length $|w| - |w'| - 2$ such that $w = w''bcw'$ holds.

Moreover, from (2) and (3), since $|aw| < |wbcw'|$ and $aw = aw''bcw'$, aw'' is a suffix of w. On the other hand, from (1') and (2'), $w'd$ is a prefix of w. Since $|w'd| + |aw''| = |w'| + |w''| + 2 = |w|$, we have $w = w'daw''$. Therefore, $w'daw'' = w''bcw'$ holds.

Claim 2. Let w', w'' be strings of constant symbols in Σ and a, b, c, d constant symbols in Σ. Then, if $b \notin \{a, d\}$ and $c \notin \{a, d\}$, then $w'daw'' \neq w''bcw'$ holds.

Proof of Claim 2. We assume that the following equation holds:

$$w'daw'' = w''bcw' \tag{1}$$

We prove this claim by an induction on $|w'| + |w''|$. W.l.o.g., we suppose that $|w'| \geq |w''|$ holds.

(i) $|w'| \geq 0$ and $|w''| = 0$: We have $w'da = bcw'$ ($b \notin \{a, d\}$ and $c \notin \{a, d\}$). It contradicts *Claim* 1.

We assume that for constant strings u and v with $|u| + |v| < |w'| + |w''|$, $vbcu \neq udav$ holds. We partition the relations between $|w'|$ and $|w''|$ into the following four parts:

(ii) $0 < |w''| \leq |w'| \leq |w''| + 2$: When either $|w'| = |w''|$ or $|w'| = |w''| + 1$, Eq. 1 is written as shown in Figs. 3, 4. Trivially, these cases contradict the conditions $b \notin \{a, d\}$ and $c \notin \{a, d\}$. When $|w'| = |w''| + 2$, Eq. 1 can be written as shown in Fig. 5. This implies $w' = w''bc = daw''$, and then contradicts *Claim* 1.

(iii) $|w''|+3 \leq |w'| \leq 2|w''|-1$: On Eq. 1, since $|w'daw''| = |w''bcw'| = |w'| + |w''| + 2$, a suffix of w' overlaps with a prefix of w' as shown in Fig. 6. I.e., there exists a constant string u of length $2|w'| - (|w'| + |w''| + 2) = |w'| - |w''| - 2$ such that u is a prefix and a suffix of w'. Since uda is of length $|w'| - |w''|$, uda is also a prefix of w'. Similarly, bcu is also a suffix of w'. Since $|w'| - (|uda| + |bcu|) = 2|w''| - |w'| \geq 1$, there exist a constant string v of length $2|w''| - |w'|$ such that $w' = udavbcu$ holds. Since w'' is a suffix of w' and $|vbcu| = (2|w''| - |w'|) + 2 + (|w'| - |w''| - 2) = |w''|$, we have $w'' = vbcu$. Similarly, we have $w'' = udav$. Thus, we have a new equation $vbcu = udav$. Since $|u| = |w'| - |w''| - 2 \leq |w''| - 3 < |w''|$ and $|v| = 2|w''| - |w'| \leq |w'|$, i.e., $|u| + |v| < |w'| + |w''|$ holds, it contradicts the induction hypothesis on $|u| + |v|$. Therefore, the claim holds.

(iv) $2|w''| \leq |w'| \leq 2|w''| + 4$: When $|w'| = 2|w''|$, we easily see that $w' = w''w''$. Therefore, $w''da = bcw''$ holds as shown in Fig. 7. It contradicts *Claim* 1. When $|w'| = 2|w''| + i$ ($i = 1, 2, 3$), Eq. 1 is written as shown in Figs. 8, 9, and 10. Trivially, these cases contradict the conditions $b \notin \{a, d\}$ and $c \notin \{a, d\}$.

(v) $2|w''| + 4 \leq |w'|$: Since the strings $w''bc$ and adw'' are a prefix and a suffix of w', respectively, and $|w''bc| + |adw''| = 2|w''| + 4$, there exists a string u with $|u| \geq 0$ such that $w' = w''bcudaw''$ holds. From Eq. 1, $w''bcudaw''daw'' = w''bcw''bcudaw''$, i.e., $udaw'' = w''bcu$ holds as shown in Fig. 12. Let $v = w''$. Since $|u| + |v| = |w'| - (2|w''| + 4) + |w''| < |w'| + |w''|$, it contradicts the induction hypothesis on $|u| + |v|$. Therefore, the claim holds.

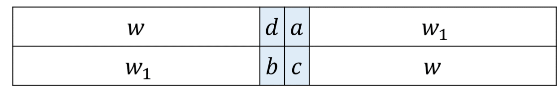

Fig. 3 $|w| = |w_1|$ における定数記号列

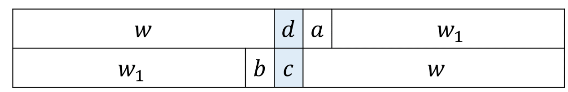

Fig. 4 $|w| = |w_1| + 1$ における定数記号列

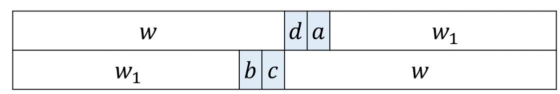

Fig. 5 $|w| = |w_1| + 2$ における定数記号列

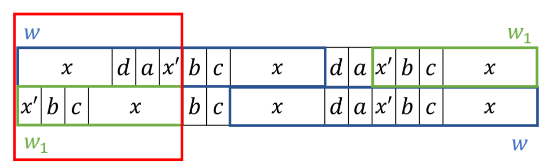

Fig. 6 $|w_1| + 3 \leq |w| \leq 2|w_1| - 1$ における定数記号列

From the above, we conclude that $w'daw'' \neq w''bcw'$ holds. (*End of Proof of Claim*)

Thus, the case of $|w| \geq |w'| + 3$ contradicts *Claim* 2.

Next, we suppose that $|w| < |w'|$ holds. Note again that from (1') and (2'), $wbcw'd$ and $w'd$ are prefixes of p_2, and from (2) and (3), $awbcw'$ and aw are suffixes of p_1.

If $|w'| = |w| + 1$, since $|wbc| = |w'd|$, we have $c = d$. This contradicts the condition $c \neq d$.

If $|w'| = |w| + 2$, since $|wbc| = |w'|$, bc is a suffix of w'. Moreover, since $w'd$ is a prefix of $wbcw'$, d is the first symbol of w'. Since aw is a suffix of w' and $|w'| = |aw| + 1$, a is the second symbol of w'. Therefore, we have $w' = wbc = daw$. This contradicts *Claim* 1.

If $|w'| \geq |w| + 3$, there exists a string w'' with $|w''| \geq 1$ such that $w' = wbcw''$ holds. Then, since $wbcw'd$ and $w'd = wbcw''d$ are prefixes of p_2, $w''d$ is a prefix of w'. Since w' and aw are suffixes of p_1 and $|w'| = |wbcw''| = |w| + |w''| + 2 > |aw|$, aw is a suffix of w'. Since $|w''d| + |aw| = |w'|$, we have $w' = w''daw$. Therefore, we have $w' = wbcw'' = w''daw$. This contradicts *Claim* 2.

Thus, the case of $A = y_1 a$, $B = bc$, and $C = dy_2$ implies the contradictions.

From the above, we conclude that if $p\{x := r\} \preceq q$ for all $r = \{ya, bc, dy\}$ ($b \notin \{a, d\}$ and $c \notin \{a, d\}$), then

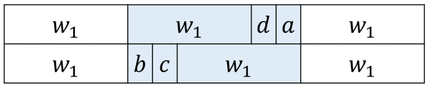

Fig. 7 $|w| = 2|w_1|$ における定数記号列

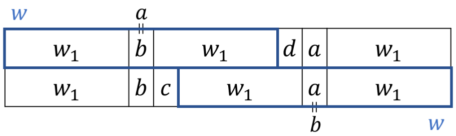

Fig. 8 $|w| = 2|w_1| + 1$ における定数記号列

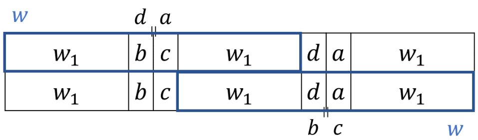

Fig. 9 $|w| = 2|w_1| + 2$ における定数記号列

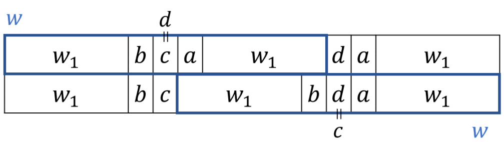

Fig. 10 $|w| = 2|w_1| + 3$ における定数記号列

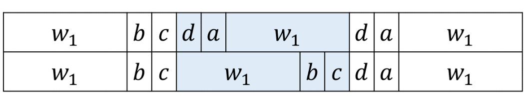

Fig. 11 $|w| = 2|w_1| + 4$ における定数記号列

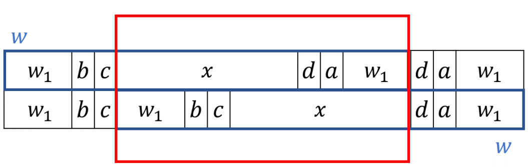

Fig. 12 $|w| \geq 2|w_1| + 5$ における定数記号列

$p\{x := xy\} \preceq q$ holds.

Lemma 7: Let Σ be an alphabet with $\Sigma \geq 3$ and p, q regular patterns on Σ. Let D be one of the following sets (i), (ii) of regular patterns on Σ. Then, if $p\{x := r\} \preceq q$ for all $r \in D$, then $p\{x := xy\} \preceq q$:

(i) $D = \{ya, bc, dy\}$ ($b = a$, $b \neq d$, and $c \notin \{a, d\}$),
(ii) $D = \{ya, bc, dy\}$ ($b \notin \{a, d\}$, $c = d$, and $c \neq a$).

Proof. It is obvious if no variable symbol appears in p. Therefore, let $p = p_1 x p_2$, where p_1, p_2 are regular patterns and x is a variable symbol. We assume that $p\{x := xy\} \npreceq q$ in order to derive the contradiction.
(i) Let $D = \{ya, bc, dy\}$ ($b = a$, $b \neq d$, and $c \notin \{a, d\}$). Since $p\{x := r\} \preceq q$ for all $r \in D$, there are three strings of length 2 corresponding to ya, bc, dy in q. Note that the three strings may appear partly overlapping. The symbols appearing in D corresponds to a variable or a constant in q. Let y_1, y_2, y_3 be variable symbols appearing in q. The strings ya and dy must correspond to the strings $y_1 a$ and dy_3 in q, respectively. There are the following three possibilities of strings in q which corresponds to bc in $p\{x := bc\}$.

$$\text{(a)}\quad bc, \quad \text{(b)}\quad y_2 c, \quad \text{(c)}\quad b y_2.$$

(a) Let A, B, C be strings where $\{A, B, C\} = \{y_1 a, bc, dy_3\}$ and let $q = q_1 A w B w' C q_2$. Since $p\{x := r\} \preceq q$ for all $r \in D$ and $p\{x := xy\} \npreceq q$ hold, the following conditions hold:

(1) $p_1 \preceq q_1$	(1′) $p_2 \preceq wBw'Cq_2$
(2) $p_1 \preceq q_1 A w$	(2′) $p_2 \preceq w'Cq_2$
(3) $p_1 \preceq q_1 A w B w'$	(3′) $p_2 \preceq q_2$

Let $q_1' = q_1 A$, $q_2' = wBw'$, $q_3' = Cq_2$. From (3) and (1′), we have $p_1 \preceq q_1' q_2'$, $p_2 \preceq q_2' q_3'$. From Lemma **??**, if q_2' contains a variable, $p \preceq q$ holds. Therefore, B must be bc. If $A = dy_3$, from (2), $p_1 \preceq q_1 dy_3 w$ holds. Let $p_1 = p_1' p_1''$, $p_1' \preceq q_1 d$, and $p_1'' \preceq y_3 w$. From (1′), we have $p = p_1 x p_2 = p_1' p_1'' x p_2 \preceq q_1 d p_1'' x w b c w' y_1 a q_2 = q\{x := p_1'' x\}$. This shows that there is a substitution θ such that $p = q\theta$ holds, and this contradicts the assumption. Therefore, we only need to consider the case where $A = y_1 a$, $B = bc$, and $C = dy_3$.

From the above, we consider two cases: one in which the symbols overlap and the other in which they do not.

(a-1) $\quad q = q_1 y_1 a w b c w' d y_3 q_2$,
(a-2) $\quad q = q_1 y_1 a c w d y_3 q_2$ ($a = b$).

(a-1) From the proof of Lemma 6, $p\{x := xy\} \preceq q$ holds. Therefore, it contradicts the assumption.

(a-2) Let $q = q_1 y_1 a c w d y_3 q_2$ ($a = b$). For this q, the following conditions hold:

(1) $p_1 \preceq q_1$	(1′) $p_2 \preceq cwdy_3q_2$
(2) $p_1 \preceq q_1 y_1$	(2′) $p_2 \preceq wdy_3q_2$
(3) $p_1 \preceq q_1 y_1 a c w d y_3$	(3′) $p_2 \preceq q_2$

If $|w| = 0$, from (1′) and (2′), the prefix of p_2 is cd and d. Therefore, $c = d$. This contradicts the fact that $c \neq d$.

If $|w| = 1$, from (1’) and (2’), the prefix of p_2 is cwd and wd. Therefore, $w = c = d$. This contradicts the fact that $c \neq d$.

If $|w| \geq 2$, then from (1′) and (2′), prefixes of p_2 is cwd and wd. Let w be $w_1 w_2 w_3 \cdots w_{n-1} w_n$ ($n \geq 2$, $w_i \in \Sigma$ for $i = 1, \ldots, n$). From $cw = wd$, a prefix of w is c and a suffix of w is d. Therefore, we have $w = c w_2 w_3 \cdots w_{n-1} d$. Since $cw = c w_2 w_3 \cdots w_{n-1} d$, $wd = c w_2 w_3 \cdots w_{n-1} d d$, from $cw = wd$, $w_i = w_{i+1}$ holds for $i = 2, \ldots, n-2$. Therefore, $c = d$. This contradicts the fact that $c \neq d$.

(b) Let $q = q_1 A w B w' C q_2$ where $\{A, B, C\} = \{y_1 a, y_2 c, d y_3\}$, and let $q = q_1 A w B w' C q_2$. Since $c \neq a$ holds, q have a substring that is corresponding to (i-2) of Lemma 4. Therefore, $p\{x := xy\} \preceq q$ holds. This contradicts the assumption.

(c) As in (b), this contradicts the assumption.
(ii) In this case, by reversing the strings p and q, we can prove that the assumption $p\{x := xy\} \preceq q$ is contradicted, as in the case of (i).

When the conditions of both Lemmas 6 and 7 are not satisfied, counterexamples exist as follows:

Proposition 1: Let Σ be an alphabet with $\sharp\Sigma \geq 3$. For a variable symbol y, let $D = \{ya, bc, dy\}$ ($b = a$ and $c = d$). There exist regular patterns p and q on Σ such that $p\{x := r\} \preceq q$ for any $r \in D$, but $p\{x := xy\} \not\preceq q$.

Proof. We give an example which shows this lemma. Let a, b, c, d, e be constant symbols in Σ and x, y, y_1, y_2 variable symbols. Let

$$p = eabcbcadabcbcadaxbcadadabcbcadade,$$
$$q = y_1abcbcadabcbcadady_2 \quad (b = a \text{ and } c = d).$$

Obviously $p\{x := xy\} \not\preceq q$ holds. For these p and q, the condition for Corollary 1 holds as follows (see also Fig. 13):

$$\begin{aligned} & p\ \{x := ya\} \\ &= (eabcbcadabcbcaday)abcadadabcbcadade \\ &= q\{y_1 := eabcbcadabcbcaday,\ y_2 := e\} \\ &\preceq q, \\ & p\ \{x := bc\} \\ &= (eabcbcad)abcbcadabcbcadad(abcbcadade) \\ &= q\{y_1 := eabcbcad,\ y_2 := abcbcadade\} \\ &\preceq q, \\ & p\ \{x := dy\} \\ &= eabcbcadabcbcadad(ybcadadabcbcadade) \\ &= q\{y_1 := e,\ y_2 := ybcadadabcbcadade\} \\ &\preceq q. \end{aligned}$$

Lemma 8: Let k be an integer with $k \geq 1$. Let Σ be an alphabet with $\sharp\Sigma = k + 2$. Let $p \in \mathcal{RP}$ in which a variable symbol x appears, and let $Q \in \mathcal{RP}^k$. If for any string $w \in \Sigma^*$ with $|w| = 2$, there exists a regular pattern $q_w \in Q$ such that $p\{x := w\} \preceq q_w$ holds, then there exists a regular pattern $q \in Q$ such that $p\{x := xy\} \preceq q$ holds, where y is a variable symbol that does not appear in q.

Proof. For any $q \in Q$, we define the sets $A(q), B(q) \subset \Sigma$ as follows:

$$A(q) = \{a \in \Sigma \mid p\{x := ay\} \preceq q,\ y \in X\},$$
$$B(q) = \{b \in \Sigma \mid p\{x := yb\} \preceq q,\ y \in X\}.$$

If there exists $q \in Q$ such that $|A(q)| \geq 2$ or $|B(q)| \geq 2$, from Lemma 4, $p\{x := xy\} \preceq q$ holds. Below, we suppose that $|A(q)| \leq 1$ and $|B(q)| \leq 1$. Let $\varnothing$ be a constant symbol that is not a member in Σ. We define the functions $\sigma_A : Q \to \Sigma \cup \{\varnothing\}$ and $\sigma_B : Q \to \Sigma \cup \{\varnothing\}$ as follows:

$$\sigma_A(q) = \begin{cases} a & \text{if } A(q) = \{a\}, \\ \varnothing & \text{if } A(q) = \emptyset. \end{cases}$$

$$\sigma_B(q) = \begin{cases} b & \text{if } B(q) = \{b\}, \\ \varnothing & \text{if } B(q) = \emptyset. \end{cases}$$

The inverse functions of σ_A and σ_B are denoted by σ_A^{-1} and σ_B^{-1}, respectively. I.e., for $a, b \in \Sigma \cup \{\varnothing\}$, let $\sigma_A^{-1}(a) = \{q \in Q \mid \sigma_A(q) = a\}$ and $\sigma_B^{-1}(b) = \{q \in Q \mid \sigma_B(q) = b\}$. A and B denotes the following subsets of Σ:

$$A = \bigcup_{q \in Q \setminus \sigma_A^{-1}(\varnothing)} A(q), \quad B = \bigcup_{q \in Q \setminus \sigma_B^{-1}(\varnothing)} B(q).$$

Then, let $A' = \Sigma \setminus A$ and $B' = \Sigma \setminus B$. For any $a, b \in \Sigma$, we use the following notations:

$$\ell_A = \sum_{a \in A} (\sharp\sigma_A^{-1}(a) - 1), \quad \ell_B = \sum_{b \in B} (\sharp\sigma_B^{-1}(b) - 1).$$

These ℓ_A and ℓ_B represent the numbers of excess duplicate symbols in A and B. We easily see the following claim:

Claim 1.

(i) $\sharp A + \sharp A' = \sharp B + \sharp B' = k + 2$,
(ii) $\sharp A + \ell_A + \sharp\sigma_A^{-1}(\varnothing) = \sharp B + \ell_B + \sharp\sigma_B^{-1}(\varnothing) = k$.

Since $\sharp\Sigma = k + 2$ and $\sharp Q = k$, $\sharp A' \geq 2$ and $\sharp B' \geq 2$ hold. We partition Q into the following subsets:

$$\begin{aligned} Q^{(\varnothing,\varnothing)} &= \sigma_A^{-1}(\varnothing) \cap \sigma_B^{-1}(\varnothing), \\ Q^{(\varnothing,\cdot)} &= \sigma_A^{-1}(\varnothing) \cap (Q \setminus \sigma_B^{-1}(\varnothing)), \\ Q^{(\cdot,\varnothing)} &= (Q \setminus \sigma_A^{-1}(\varnothing)) \cap \sigma_B^{-1}(\varnothing), \\ Q^{(\cdot,\cdot)} &= (Q \setminus \sigma_A^{-1}(\varnothing)) \cap (Q \setminus \sigma_B^{-1}(\varnothing)). \end{aligned}$$

From the condition of this lemma, for any string $w \in \Sigma^*$ with $|w| = 2$, there exists a regular pattern $q_w \in Q$ such that $p\{x := w\} \preceq q_w$ holds. It is easy to see that if $w \in (A \cdot B) \cup (A' \cdot B) \cup (A \cdot B')$, there exists a regular pattern $q_w \in Q^{(\varnothing,\cdot)} \cup Q^{(\cdot,\varnothing)} \cup Q^{(\cdot,\cdot)}$ such that $p\{x := w\} \preceq q_w$ holds. For $w = a'b' \in A' \cdot B'$, we must have $q_w \in Q$ that satisfies that $p\{x := w\} \preceq q_w$. The following two claims are proven from Lemmas 4 and 5:

Claim 2. If there exist $q \in Q^{(\varnothing,\varnothing)}$ and distinct 5 strings $w_i \in A' \cdot B'$ ($1 \leq i \leq 5$) such that $p\{x := w_i\} \preceq q$ holds ($1 \leq i \leq 5$), then $p\{x := xy\} \preceq q$ holds.

Claim 3. If there exist $q \in Q^{(\varnothing,\cdot)} \cup Q^{(\cdot,\varnothing)}$ and distinct 3 strings $w_i \in A' \cdot B'$ ($1 \leq i \leq 3$) such that $p\{x := w_i\} \preceq q$ holds ($1 \leq i \leq 3$), then $p\{x := xy\} \preceq q$ holds.

If there exist a regular pattern $q \in Q^{(\varnothing,\varnothing)} \cup Q^{(\varnothing,\cdot)} \cup Q^{(\cdot,\varnothing)}$ and enough strings $w \in A' \cdot B'$ such that either of the conditions of *Claims* 2 and 3 is satisfied, this lemma holds. Then, we assume that it is not the case.

Assumption 1. There is no regular pattern $q \in Q^{(\varnothing,\varnothing)}$ and 5 strings $w \in A' \cdot B'$ such that the condition of *Claim* 2 is satisfied and there is no regular pattern $q \in Q^{(\varnothing,\cdot)} \cup Q^{(\cdot,\varnothing)}$ and 3 strings $w \in A' \cdot B'$ such that the condition of *Claim* 3 is satisfied.

Let $\mathcal{L}_1 = \sharp\{w \in A' \cdot B' \mid \exists q \in Q^{(\varnothing,\varnothing)} \cup Q^{(\varnothing,\cdot)} \cup$

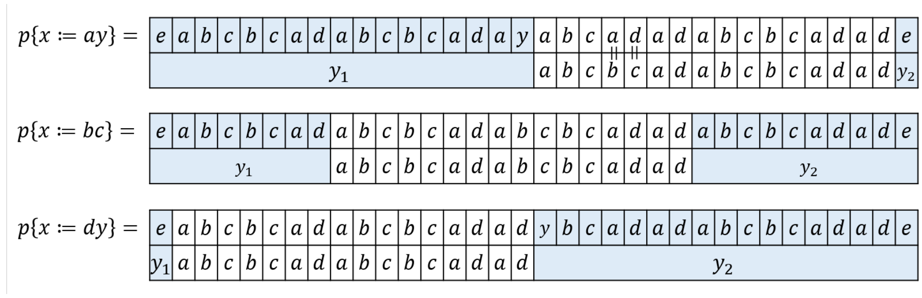

Fig. 13 Substitutions for p and each correspondence to q.

$Q^{(\cdot,\emptyset)}$ s.t. $p\{x := w\} \preceq q\}$. Then, by *Claim* 1,

$$\begin{aligned}\mathcal{L}_1 &\le 4\sharp Q^{(\emptyset,\emptyset)} + 2\sharp Q^{(\emptyset,\cdot)} + 2\sharp Q^{(\cdot,\emptyset)}\\ &= 2(Q^{(\emptyset,\emptyset)} + \sharp Q^{(\emptyset,\cdot)}) + 2(Q^{(\emptyset,\emptyset)} + \sharp Q^{(\cdot,\emptyset)})\\ &= 2\sharp\sigma_A^{-1}(\emptyset) + 2\sharp\sigma_B^{-1}(\emptyset)\\ &= 2(k - \sharp A - \ell_A) + 2(k - \sharp B - \ell_B)\\ &= 2(\sharp A' - \ell_A - 2) + 2(\sharp B' - \ell_B - 2)\\ &= 2(\sharp A' + \sharp B') - 2(\ell_A + \ell_B) - 8.\end{aligned}$$

Next, we partition $Q^{(\cdot,\cdot)}$ into the following two subsets:

$$\begin{aligned}Q_1^{(\cdot,\cdot)} &= \{q \in Q^{(\cdot,\cdot)} \mid \sigma_A(q) \in B \text{ or } \sigma_B(q) \in A\},\\ Q_2^{(\cdot,\cdot)} &= \{q \in Q^{(\cdot,\cdot)} \mid \sigma_A(q) \in B' \text{ and } \sigma_B(q) \in A'\}.\end{aligned}$$

We show the next two claims on $Q_1^{(\cdot,\cdot)}$ and $Q_2^{(\cdot,\cdot)}$:

Claim 4. If there exist $q \in Q_1^{(\cdot,\cdot)}$ and a string $a'b' \in A' \cdot B'$ such that $p\{x := a'b'\} \preceq q$ holds, then $p\{x := xy\} \preceq q$ holds.

Proof of Claim 4. Suppose that both $\sigma_A(q) \in B$ and $\sigma_B(q) \in A$ hold. Then, since $a' \notin \{\sigma_A(q), \sigma_B(q)\} \subset A \cap B$ and $b' \notin \{\sigma_A(q), \sigma_B(q)\} \subset A \cap B$, from Lemma 6, $p\{x := xy\} \preceq q$ holds. Suppose that $\sigma_A(q) \in B$ and $\sigma_B(q) \in A'$. If $a' = \sigma_B(q)$, since $a' \in B$, $a' \neq b'$ holds. Since $\sigma_A(q) \in B$, $b' \neq \sigma_A(q)$ holds. I.e., $a' = \sigma_B(q)$, $a' \neq \sigma_A(q)$, and $b' \notin \{\sigma_A(q), \sigma_B(q)\}$ hold. Therefore, from Lemma 7, $p\{x := xy\} \preceq q$ holds. If $a' \neq \sigma_B(q)$, since $b' \neq \sigma_A(q)$, from Lemma 6, $p\{x := xy\} \preceq q$ holds. Similarly, the case that $\sigma_A(q) \in B'$ and $\sigma_B(q) \in A$ is proven. (*End of Proof of Claim*)

Claim 5. If there exist $q \in Q_2^{(\cdot,\cdot)}$ and a string $a'b' \in A' \cdot B'$ such that ($a' \neq \sigma_B(q)$ or $b' \neq \sigma_A(q)$) and $p\{x := a'b'\} \preceq q$ hold, then $p\{x := xy\} \preceq q$ holds.

Proof of Claim 5. When $a' = b'$, since $\sigma_A(q) \neq \sigma_B(q)$, from Lemma 6, this claim holds. Similarly, when $a' \neq b'$, from Lemma 6 or Lemma 7, this holds. (*End of Proof of Claim*)

If there exist a regular pattern $q \in Q_2^{(\cdot,\cdot)}$ and a string $w \in A' \cdot B'$ such that the condition of *Claim* 5 is satisfied, this lemma holds. Then, we also assume that it is not the case.

Assumption 2. There is no $q \in Q_2^{(\cdot,\cdot)}$ and a string $a'b' \in A' \cdot B'$ such that the condition of *Claim* 5 is satisfied.

Let $\mathcal{L}_2 = \sharp\{a'b' \in A' \cdot B' \mid \exists q \in Q_2^{(\cdot,\cdot)}$ s.t. $p\{x := a'b'\} \preceq q\}$. For any $a'b' \in A' \cdot B'$ and $q \in Q_2^{(\cdot,\cdot)}$, if $a' = \sigma_B(q)$ and $b' = \sigma_A(q)$ hold (it is the condition of Proposition 1), by considering the duplicate numbers ℓ_A and ℓ_B, we have the following inequality:

$$\mathcal{L}_2 \le \min\{\sharp A' + \ell_B, \sharp B' + \ell_A\}.$$

We show the last claim:

Claim 6. $\sharp A' \times \sharp B' - \mathcal{L}_1 - \mathcal{L}_2 \ge 2$.

Proof of Claim 6. First we prove the inequality when $\sharp A \le k-1$ and $\sharp B \le k-1$, i.e., $\sharp A' \ge 3$ and $\sharp B' \ge 3$ hold. Since $\mathcal{L}_2 \le \frac{1}{2}(\sharp A' + \sharp B' + \ell_A + \ell_B)$,

$$\begin{aligned}&\sharp A' \times \sharp B' - \mathcal{L}_1 - \mathcal{L}_2\\ \ge\ &\sharp A' \times \sharp B' - (2(\sharp A' + \sharp B') - 2(\ell_A + \ell_B) - 8)\\ &- \frac{1}{2}(\sharp A' + \sharp B' + \ell_A + \ell_B)\\ =\ &\sharp A' \times \sharp B' - \frac{5}{2}(\sharp A' + \sharp B') + \frac{3}{2}(\ell_A + \ell_B) + 8\\ =&(\sharp A' - \frac{5}{2})(\sharp B' - \frac{5}{2}) + \frac{3}{2}(\ell_A + \ell_B) + \frac{7}{4} \ge 2.\end{aligned}$$

When $\sharp A = k$ and $\sharp B \le k$, i.e., $\sharp A' = 2$ and $\sharp B' \ge 2$ hold, since $\ell_A = 0$, $\mathcal{L}_1 \le 2\sharp B' - 2\ell_B - 4$ holds. Moreover, $\mathcal{L}_2 \le \min\{\sharp B', \ell_B + 2\}$ holds. From *Claim* 1, $\ell_B + 2 = k - \sharp\sigma_B^{-1}(\emptyset) - \sharp B = \sharp B' - \sharp\sigma_B^{-1}(\emptyset)$ hold. Therefore, $\mathcal{L}_2 \le \ell_B + 2$ holds. Thus,

$$\begin{aligned}&\sharp A' \times \sharp B' - \mathcal{L}_1 - \mathcal{L}_2\\ \ge\ &2\sharp B' - (2\sharp B' - 2\ell_B - 4) - (\ell_B + 2)\\ =&\ell_B + 2 \ge 2.\end{aligned}$$

Similarly, the case when $\sharp A \le k$ and $\sharp B = k$ is proven. (*End of Proof of Claim*)

Under *Assumptions* 1 and 2, from *Claim* 6, there exist at least two $w \in A' \cdot B'$ and a regular pattern $q \in Q_1^{(\cdot,\cdot)}$ such that the condition of *Claim* 4 is satisfied. Therefore, for such a regular pattern q, $p\{x := xy\} \preceq q$ holds.

Lemma 9 (Sato et al.[1]): Let Σ be a finite alphabet with $\sharp\Sigma \ge 3$ and p, q be regular patterns. If there exists a constant symbol $a \in \Sigma$ such that $p\{x := a\} \preceq q$ and $p\{x := xy\} \preceq q$, then $p \preceq q$ holds, where y is a variable symbol that does not appear in q.

From the Lemma8 and Lemma9, we have the following theorem.

Theorem 4: Let $k \geq, \sharp\Sigma \geq 2k-1$, $P \in \mathcal{RP}^+$ and $Q \in \mathcal{RP}^k$. Then, the following (i),(ii) and (iii) are equivalent:

$$\text{(i) } S_2(P) \subseteq L(Q), \text{ (ii) } P \sqsubseteq Q, \text{ (iii) } L(P) \subseteq L(Q).$$

Proof. it is clear that (ii) implies (iii) and (iii) implies (i). From Theorem3, if $\sharp\Sigma \geq 2k+1$, then (i) implies (ii). Let $\sharp Q = k$, $p \in P$, $\sharp\Sigma = 2k-1$ or $2k$. Then, we show that (i) implies (ii). It suffices to show that $S_2(p) \subseteq L(Q)$ implies $P \sqsubseteq Q$ for any regular pattern $p \in \mathcal{RP}$. The proof is done by mathematical induction on n, where n is the number of variable symbols appears in p.

In case $n = 0$, $S_2(p) = \{p\}$. By (i), we have $\{p\} = L(Q)$. Thus, $p \preceq q$ for some $q \in Q$.

For $n \geq 0$, we assume that it is valid for any regular pattern p with n variable symbols. Let p be a regular pattern such that $n+1$ variable symbols appear in p and $S_2(p) \subseteq L(Q)$.

We assume that $p \not\sqsubseteq Q$, that is, $p \not\preceq q_i$ for any $i \in \{1,\ldots,k\}$. Let $Q = \{q_1,\ldots,q_k\}$ and p_1, p_2 regular patterns, x a variable symbol with $p = p_1xp_2$. For $a,b \in \Sigma$, let $p_a = p\{x := a\}$ and $p_{ab} = p\{x := ab\}$. Both p_a and p_{ab} have n variable symbols respectively. Thus, $S_2(p_a) \subseteq L(Q)$ and $S_2(p_{ab}) \subseteq L(P)$ hold. By the induction hypothesis, there exist $i, i' \in \{1,\ldots,k\}$ such that $p_a \preceq q_i$ and $p_{ab} \preceq q_{i'}$. Let $D_i = \{a \in \Sigma \mid p\{x := a\} \preceq q_i\}$ $(i = 1,\ldots,k)$. We assume that $\sharp D_i \geq 3$ for some $i \in \{1,\ldots,k\}$. By Lemma **??**, we have $p \preceq q_i$. This contradicts the assumption. Thus, we have $\sharp D_i \leq 2$ for any $i \in \{1,\ldots,k\}$. If $\sharp\Sigma = 2k-1$, then $\sharp D_i = 2$ or $\sharp D_i = 1$ for any $i \in \{1,\ldots,k\}$. Moreover, If $\sharp\Sigma = 2k$, then $\sharp D_i = 2$ for any $i \in \{1,\ldots,k\}$. Since $k \geq 3$, $2k+1 \geq k+2$ holds. By Lemma 8, there exists $i \in \{1,\ldots,k\}$ such that $p\{x := xy\} \preceq q_i$. Therefore, by Lemma 9, we have $p \preceq q_i$. This contradicts the assumption. Thus, (i) implies (ii).

From Theorem 4, the following corollary holds.

Corollary 2: Let $k \geq 3$, $\sharp\Sigma \geq 2k-1$ and $P \in \mathcal{RP}^+$. Then, $S_2(P)$ is a characteristic set for $L(P)$ within $\mathcal{RPL}^k$.

Lemma 10 (Sato et al.[1]): Let $k \geq 3$ and $\sharp\Sigma \leq 2k-2$. Then, $\mathcal{RP}^k$ does not have compactness with respect to containment.

Proof. Let $\Sigma = \{a_1,\ldots,a_{k-1},b_1,\ldots,b_{k-1}\}$ and p, q_i regular patterns, $w_i \in \Sigma^*$ $(i = 1,\ldots,k-1)$ defined in a similar way to Example **??**. Let $q_k = x_1a_1w_1xyw_1b_1x_2$. Since $p\{x := a_i\} = x_1a_1w_1a_iw_1b_1x_2 \preceq q_i$ and $p\{x := b_i\} = x_1a_1w_1b_iw_1b_1x_2 \preceq q_i$ for any $i \in \{1,\ldots,k-1\}$, we have $S_1(p) \subseteq \bigcup_{i=1}^{k-1} L(q_i)$. For any $w \in \{s \in \Sigma^+ \mid |s| \geq 2\}$, $p\{x := w\} = x_1a_1w_1ww_1b_1x_2 \preceq q_k$. Thus, we have $L(p) \subseteq L(Q)$. By Theorem 1, since $p \not\preceq q_i$, $L(p) \not\subseteq L(q_i)$ for any $i \in \{1,\ldots,k\}$. Therefore, $\mathcal{RP}^k$ does not have compactness with respect to containment.

From Theorem 4 and Lemma 10, we have the following thorem.

Theorem 5: Let $k \geq 3$ and $\sharp\Sigma \geq 2k-1$. Then, $\mathcal{RP}^k$ has compactness with respect to containment.

In case $k = 2$, we have the following theorem.

Theorem 6: Let $\sharp\Sigma \geq 4$, $P \in \mathcal{RP}^+$ and $Q \in \mathcal{RP}^2$. The following (i), (ii) and (iii) are equivalent:

$$\text{(i) } S_2(P) \subseteq L(Q), \quad \text{(ii) } P \sqsubseteq Q, \quad \text{(iii) } L(P) \subseteq L(Q).$$

Proof. It is clear that (ii) implies (iii), and (iii) implies (i). Thus, we show that (i) implies (ii). It suffices to show that $S_2(p) \subseteq L(Q)$ implies $P \sqsubseteq Q$ for any regular pattern $p \in \mathcal{RP}$. Let $Q = \{q_1, q_2\}$. The proof is done by mathematical induction on n, where n is the number of variable symbols appearing in p. In case $n = 0$, $p \in \Sigma^+$. Since $S_2(p) = \{p\} \subseteq L(Q)$, we have $p \preceq q$ for some $q \in Q$. For $n \geq 0$, we assume that it is valid for any regular pattern p with n variable symbols. Let p be a regular pattern such that $n+1$ variable symbols appear in p, ans $S_2(p) \subseteq L(Q)$. We assume that $p \not\preceq q_i$ $(i = 1,2)$. Let p_1, p_2 be regular patterns and x a variable symbol with $p = p_1xp_2$. For $a,b \in \Sigma$, let $p_a = p\{x := a\}$ and $p_{ab} = p\{x := ab\}$. Note that p_a and p_{ab} have n variable symbols. Thus, by the assumption, $S_2(p_a) \subseteq L(Q)$ and $S_2(p_{ab}) \subseteq L(Q)$ implies $p_a \preceq q_i$ and $p_{ab} \preceq q_{i'}$ for some $i, i' \in \{1,2\}$. Let $D_i = \{a \in \Sigma \mid p\{x := a\} \preceq q_i\}$ $(i = 1,2)$. By Lemma **??**, if $\sharp D_i \geq 3$ for some $i \in \{1,2\}$, then $p \preceq q_i$. This contradicts that $p \not\preceq q_i$ $(i = 1,2)$. Thus, we have $\sharp D_i \leq 2$ for any $i \in \{1,2\}$. Since $\sharp\Sigma \geq 4$, We consider that $\sharp D_1 = 2$ and $\sharp D_2 = 2$. From Lemma 8, $p\{x := xy\} \preceq q_i$ for some $i \in \{1,2\}$. From Lemma 9, we have $p \preceq q_i$ for some $i \in \{1,2\}$. This contradicts that $p \not\preceq q_i$ $(i = 1,2)$. Therefore, (i) implies (ii).

The next example is a counter-example of Theorem 6.

Example 1: Let $\Sigma = \{a,b,c\}$, p, q_1, q_2 regular patterns and x, x', x'' variable symbols such that $p = x'axbx''$, $q_1 = x'abx''$ and $q_2 = x'cx''$. Let $w \in \Sigma^+$. If w contains c, then $p\{x := w\} \preceq q_2$. On the other hand, if w does not contain c, then $p\{x := w\} \preceq q_1$. Thus, $L(p) \subseteq L(q_1) \cup L(q_2)$. However, $p \not\preceq q_1$ and $p \not\preceq q_2$.

From Theorem 6, we have that following two corollaries.

Corollary 3: Let $\sharp\Sigma \geq 4$ and $P \in \mathcal{RP}^+$. Then, $S_2(P)$ is a characteristic set for $L(P)$ within $\mathcal{RPL}^2$.

Corollary 4: Let $\sharp\Sigma \geq 4$. Then, $\mathcal{RP}^2$ has compactness with respect to containment.

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．例えば，パターン $axybc$ は正規パターンであるが，非隣接変数正規パターンではない．パターン $axbcy$ は非隣接変数正規パターンである．

$\mathcal{RP}_{NAV}$ を非隣接変数正規パターン全体の集合とする．$\mathcal{RP}_{NAV}$ の空でない有限部分集合の集合を $\mathcal{RP}^{+}_{NAV}$ で，高々k ($k \geq 1$) 個のパターンから成る $\mathcal{RP}_{NAV}$ の部分集合 $\{P \in \mathcal{RP}^{+}_{NAV} \mid \sharp P \leq k\}$ を $\mathcal{RP}^{k}_{NAV}$ で表す．このとき，次の定理が成り立つ．

Theorem 7: $\sharp\Sigma \geq k+2$, $P \in \mathcal{RP}^{+}_{NAV}$, $Q \in \mathcal{RP}^{k}_{NAV}$ とする．このとき，以下の (i), (ii), (iii) は同値である．

$$\text{(i) } S_2(P) \subseteq L(Q), \text{ (ii) } P \sqsubseteq Q, \text{ (iii) } L(P) \subseteq L(Q).$$

Proof. 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任意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．$p \npreceq q_i$ ($i = 1, 2$) と仮定する．非隣接変数正規パターン p を $p = p_1xp_2, Q = \{q_1, \ldots, q_k\}$ 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，任意の i ($i = 1, \ldots, k$) 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，任意の $a, b \in \Sigma$ 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 8 より，ある i に対して $p\{x := xy\} \preceq q_i$ が成り立つ．このとき，$p\{x := xy\} = p_1xyp_2$ 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．

Corollary 5: $\sharp\Sigma \geq k+2$，$P \in \mathcal{RP}^{+}_{NAV}$ とする．このとき，$S_2(P)$ は $\mathcal{RPL}^{k}_{\mathcal{NAV}}$ における $L(P)$ の特徴集合である．

Lemma 11: $\sharp\Sigma \leq k+1$ とする．このとき，$\mathcal{RP}^{k}_{NAV}$ は包含に関してコンパクト性を持たない．

Proof. $\Sigma = \{a_1, \ldots, a_{k+1}\}$ を $k+1$ 個の定数記号から成る集合，p, q_i を正規パターンとする．$p\{x := a_iy\} \preceq q_i$ かつ $p\{x := ya_{i+1}\} \preceq q_i$ ($i = 1, 2, \ldots, k$) とする．$p\{x := a_{k+1}a_1\} \preceq q_1$ であるとき，$S_2(p) \backslash S_1(p) \subseteq \bigcup_{i=1}^{k} L(q_i)$ となる．すなわち，$L(p) \subseteq L(Q)$ である．しかし，$p \npreceq q_i$ であるため，$L(p) \nsubseteq L(q_i)$ ($i = 1, 2, \ldots k$) である．したがって，$\mathcal{RP}^{k}_{NAV}$ は包含に関するコンパクト性を持たない．

コンパクト性をもたない例を例 2 に示す．

定理 7 と補題 11 より，次の定理が成り立つ．

Theorem 8: $\sharp\Sigma \geq k+2$ 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．

5. おわりに

本稿では，高々k ($k \geq 2$) 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 $\mathcal{RP}^{k}_{NAV}$ 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 $k+2$ 以上のとき，$\mathcal{RP}^{k}_{NAV}$ 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 [4] 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，$\mathcal{RP}^{k}_{NAV}$ 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 [5] の有限和や正則 FGS 言語 [6] に拡張することが考えられる．

謝辞

本研究は JSPS 科研費 19K12103, 20K04973, 21K12021, 22K12172 の助成を受けたものである．

References

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

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

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

[4] 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, Springer LNCS 775, pp.649-660, 1994.

[5] Y. Suzuki, T. Shoudai, T. Uchida and T. Miyahara, Ordered Term Tree Languages Which are Polynomial Time Inductively Inferable from Positive Data, Theoretical Computer Science, 350(1):63-90, 2006.

[6] T. Uchida, T. Shoudai, S. Miyano, Parallel Algorithms for Refutation Tree Problem on Formal Graph Systems, IEICE Trans. Inf. & Syst., E78-D(2):99-112, 1995.

Example 2: $\Sigma = \{a_1, a_2, a_3, a_4\}$ を 4 つの定数記号から成る集合，p, q_1, q_2, q_3 を正規パターン，x, x', x'' を変数記号とする．p, q_1, q_2, q_3 を以下のように定義する．

$$\begin{aligned}
p &= x' a_3 a_1 a_4 a_1 a_4 a_1 a_1 a_4 a_1 a_3 a_2 a_1 a_4 a_1 a_4 a_1 a_1 a_4 a_1 x a_1 a_4 a_1 a_4 a_1 a_1 a_4 a_1 a_3 a_2 a_1 a_4 a_1 a_4 a_1 a_1 a_4 a_1 a_2 x'', \\
q_1 &= x' a_3 a_1 a_4 a_1 a_4 a_1 a_1 a_4 a_1 a_3 a_2 a_1 a_4 a_1 a_4 a_1 a_1 a_4 a_1 a_2 x'', \\
q_2 &= x' a_2 a_1 a_4 a_1 a_4 a_1 a_1 a_4 a_1 a_3 x'', \\
q_3 &= x' a_1 a_1 a_4 a_1 a_4 x''.
\end{aligned}$$

これは，$L(p) \subseteq L(q_1) \cup L(q_2) \cup L(q_3)$ となる．しかし，$p \npreceq q_1$，$p \npreceq q_2$ かつ $p \npreceq q_3$ である．