社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－AL－117 (11)
2008／3／7

二値アルファベット上の有限オートマトンの等価変換と状態数解析

松浦 昭洋†, 齋藤 祐輔†

†東京電機大学 理工学部 情報システム工学科

概 要

任意の非負整数 α に対して, n 状態の非決定性有限オートマトン (NFA) で, その等価な最小決定性有限オートマトン (DFA) が $2^n-\alpha$ 状態を持つものが存在するか, という問題は Iwama ら [3] によって提起された. 本稿では, $n \geq 11$, $\alpha \leq 3n-3$, $\lfloor(\alpha-1)/3\rfloor$ が奇数でかつ n と互いに素である, という条件を満たす n と α に対して, n 状態 NFA で, その等価な最小 DFA が $2^n-\alpha$ 状態を持つものが存在することを示す. また, $\alpha \leq 4n-5$ の範囲の 4 の倍数でない α に対しても, 同様の NFA の存在を示す.

Equivalent Transformation of Finite Automata over a Two-Letter Alphabet and the State Complexity

Akihiro MATSUURA†, Yusuke SAITO†

†Department of Computers and Systems Engineering
School of Science and Engineering
Tokyo Denki University

概 要

In [3], Iwama et al. posed a problem which asks whether, for a given integer α, there exists a minimum n-state nondeterministic finite automaton (NFA) that is equivalent to a minimum deterministic finite automaton (DFA) with exactly $2^n-\alpha$ states. In this paper, we show that for all integers $n \geq 11$ and α, such that $\alpha \leq 3n-3$ and satisfying that $\lfloor(\alpha-1)/3\rfloor$ is odd and is relatively prime with n, there exists a minimum n-state NFA whose equivalent minimum DFA has exactly $2^n-\alpha$ states. The results are then extended for α such that $\alpha \leq 4n-5$ when α is not divisible by four.

1 はじめに

有限オートマトンは, 基本的な計算モデルであり, コンパイラにおける言語解析や正規表現, ソフトウェアの設計や動作解析といった応用面でも重要な役割を果たす概念である. 決定性有限オートマトン (Deterministic Finite Automata, DFA), 非決定性有限オートマトン (Nondeterministic Finite Automata, NFA) をはじめ, テープや遷移の種類によって様々なモデルの研究が行われている. NFA は遷移の非決定性ゆえ, 動作を解析したりプログラミングすることが困難であり, そのため等価な DFA に変換される. Rabin と Scott による "subset construction" により, どのような NFA も等価な DFA に変換可能である [7]. NFA の状態数を n とするとき, 変換後の DFA の状態数は 2^n まで増加することがよく知られるが, 等価な DFA の状態数として, どのような値が実現可能かは, 最近までよく分かっていなかった.

n 状態 NFA と等価な最小 DFA の状態数を $2^n-\alpha$ としたとき, どのような α に対して, そのような状態数の関係を持つ NFA と DFA が存在するかは Iwama ら [3] によって問題として提起され, 入力アルファベット (遷移記号) が二値である場合, $\alpha = 2^k, 2^k+1$ $(k \leq n/2-2)$

のとき, 等価最小の DFA が $2^n-\alpha$ 状態を持つような n 状態 NFA が示された. 以来, 本問題は, 近年の有限オートマトンの状態数解析への注目もあり, 有限オートマトンの研究者の関心を集めている. [4] において, 上記のような NFA と DFA の組が存在しない α は "マジックナンバー" と呼ばれ, 入力アルファベットが二値である場合, $\alpha \leq 2n-2$ の範囲にある全ての自然数 (但し, n と α はある素条件を満たす) がマジックナンバーでない, すなわち, その範囲の α に対して, 等価最小の DFA が $2^n-\alpha$ 状態持つような n 状態 NFA の存在が示された.

アルファベットが unary である場合, Chrobak により, どのような n 状態 NFA も $O(e^{\sqrt{n \ln n}})$ 状態の最小 DFA と等価であることが示され [1], さらに最近, 上限内の殆どの数 α がマジックナンバーであることが Geffert によって示された [2]. 一方で, アルファベットのサイズが $2n$ である場合, $0 \leq \alpha \leq 2^n-n$ を満たす任意の α がマジックナンバーでないことが Jirásková によって示された [6]. 彼らは, アルファベットのサイズが 4 以上のときにも, $0 \leq \alpha \leq 2^n-n$ を満たす任意の α がマジックナンバーでないことを示した [5]. また, Zijl は, マジックナンバーに関する考察を Symmetric Difference NFA と呼ばれる有限オートマトンに対して行った [8].

本稿では, アルファベットが二値の場合に, マジックナンバーでない α の範囲を一定の条件の下, $4n$ 程度まで伸長する. 正確には, まず, $n \geq 11$, $28 \leq \alpha \leq 3n-3$, $\lfloor(\alpha-1)/3\rfloor$ が奇数で n と互いに素である, という条件を満たす n と α に対して, 等価最小の DFA が $2^n-\alpha$ 状態持つような n 状態 NFA の存在を示す. また, $37 \leq \alpha \leq 4n-5$ を満たす 4 の倍数でない α に対しても同様の NFA の存在を示す.

2 準備

決定性有限オートマトン A を以下の五つ組によって定義する.

A = $(Q, \Sigma, q_0, F, \delta)$
Q : 状態集合
Σ : 遷移記号の集合
q_0 : 初期状態
F : 受理状態の集合
δ : $Q \times \Sigma \to Q$: 遷移関数

非決定性有限オートマトンは, 五つ組 $B=(Q, \Sigma, Q_0, F, \delta)$ であり, Q, Σ, F は DFA と同様の定義され, Q_0 は初期状態の集合であり, 遷移関数は次のように定義される.

$$\delta : Q \times \Sigma \to 2^Q$$

但し, 2^Q は Q のベキ集合を表す. 本稿では, $\Sigma=\{0,1\}$ とする.

NFA B が与えられたとき, B と等価な DFA を構成する手法である "subset construction" [7] を確認する. subset construction により, $B=(Q, \Sigma, Q_0, F, \delta)$ と等価な DFA A は以下のように構成される. A の各状態は, B の状態集合 Q の部分集合から成り, Q の部分集合 $\{p_{i_0}, p_{i_1}, \cdots, p_{i_k}\}$ に対応する A の状態を $[p_{i_0}, p_{i_1}, \cdots, p_{i_k}]$ と表す. 初期状態は $q_0=[q_0]$ であり, A における遷移関数 δ' : $2^Q \times \Sigma \to 2^Q$ は, B の遷移関数 δ から次のように定義される. $[p_{i_0}, p_{i_1}, \cdots, p_{i_k}]$, $[p_{j_0}, p_{j_1}, \cdots, p_{j_l}] \in 2^Q$, $\sigma \in \Sigma$ に対して,

$$\delta'([p_{i_0}, p_{i_1}, \cdots, p_{i_k}], \sigma) = [p_{j_0}, p_{j_1}, \cdots, p_{j_l}]$$

$$\text{iff} \quad \bigcup_{s=0}^{k} \delta(p_{i_s}, \sigma) = \{p_{j_0}, p_{j_1}, \cdots, p_{j_l}\}.$$

一般に, DFA において, 二つの状態 P_1 と P_2 が同値であるとは, その遷移関数を δ としたとき, 任意の $x \in \Sigma^*$ に対して, $\delta(P_1, x) \in F$ iff $\delta(P_2, x) \in F$ が成り立つときに言う. 初期状態から到達可能な全ての状態が互いに同値でないならば, その DFA は最小となる [7].

3 主要結果

以下の二定理が本稿の主要結果である.

定理 1 $n \geq 11$, $28 \leq \alpha \leq 3n-3$, $\lfloor(\alpha-1)/3\rfloor$ が奇数で, かつ n と互いに素である, という条件を満たす任意の n と α に対して, 等価最小の DFA の状態数が $2^n-\alpha$ である n 状態 NFA が存在する.

定理 2 $n \geq 11$, $37 \leq \alpha \leq 4n-5$, α が 4 の倍数でなく, $\lfloor(\alpha-1)/4\rfloor$ が n と互いに素であ

る, という条件を満たす任意の n と α に対して, 等価最小 DFA の状態数が $2^n-\alpha$ である n 状態 NFA が存在する.

定理1および定理2により, 上記範囲の α がマジックナンバーでないことが示される.

4 定理1の証明

4.1 NFA の構成

k と m を互いに素な自然数とし, $k \geq 9$, $m \geq 2$ で, かつ k は奇数とする.

まず, $\alpha = 3k+1$ を満たす NFA M_1 を以下の五つ組で定義する.

$$
\begin{aligned}
M_1 &= (T \cup S, \Sigma, Q_0, \delta, F) \\
T &= \{t_0, t_1, \cdots, t_{k-1}\} \\
S &= \{s_0, s_1 \cdots, s_{m-1}\} \\
Q_0 &= F = \{t_0\}
\end{aligned}
$$

$$
\delta(t_i,\sigma) = \begin{cases} t_{i+1} & (0 \leq i \leq k-1, \sigma = 0) \\ t_i & (i \neq 3, \sigma = 1) \\ s_0 & (i = 3, \sigma = 1) \\ t_0 & (4 \leq i \leq k-4, \sigma = 1) \\ t_2 & (i = 4, \sigma = 1) \end{cases}
$$

$$
\delta(s_j,\sigma) = \begin{cases} s_{j+1} & (0 \leq j \leq m-1, \sigma = 0) \\ s_1 & (j = 0, \sigma = 1) \\ t_0, t_3 & (j = 1, \sigma = 1) \\ s_j & (2 \leq j \leq m-1, \sigma = 1) \end{cases}
$$

なお, 以後, 遷移関数 δ における $i+1$ $j+1$ など, 添字の加減算は, T に関しては mod k で, S に関しては mod m で考える. M_1 の全状態数は $k+m=n$ である.

0 (または 1) を一回以上連続して読む遷移を 0 遷移 (1 遷移) と呼ぶことにする. また, NFA M_1 と等価かつ最小の DFA M_1' の状態 P において, T の状態から成る部分, すなわち $P \cap T$ を P_T と表し, P の T 状態と呼ぶ. 同様に, S の状態から成る部分を $P \cap S$ を P_S と表し, P の S 状態と呼ぶ. 状態 P に含まれるの状態数を $|P|$ と表す.

4.2 DFA の各状態への到達可能性

NFA M_1 と等価な最小 DFA M_1' の状態数を求めるため, 以下の補題群を示す.

補題1 $|P_S| = 0$, かつ $0 \leq |P_T| \leq 3$ を満たす状態 P について, 以下が成り立つ.

1. $|P_T| = 0$ のとき, すなわち空集合は到達不可能である.
2. $|P_T| = 1$ となる状態 P は全て到達可能である.
3. $|P_T| = 2$ となる状態 P は, $[t_i, t_{i+1}]$, $[t_i, t_{i+2}]$ と表せる $2k$ 個の状態のみ到達不可能で, 残りは全て到達可能である.
4. $|P_T| = 3$ かつ $\{t_i, t_{i+3}\} \subset P$ でない状態 P は, $[t_i, t_{i+1}, t_{i+2}]$ と表せる k 個の状態のみ到達不可能で, 残りは全て到達可能である.

補題1の4に関して, $|P_T| = 3$ かつ $\{t_i, t_{i+3}\} \subset P$ を満たす状態 P は, 以下の補題2の中で考慮される.

補題2 $|P_S| = 0$, $|P_T| \geq 3$, かつ $\{t_i, t_{i+3}\} \subset P$ を満たす状態 P は, $(|Q_S|, |Q_T|) = (0, |P_T|-1)$ または $(1, |P_T|-2)$ を満たす状態 Q から到達可能である.

補題3 $|P_S| = 0$, $|P_T| \geq 4$, かつ $\{t_i, t_{i+3}\} \subset P$ でない状態 P は, $(|Q_S|, |Q_T|) = (0, |P_T|-1)$ を満たす状態 Q から到達可能である.

補題4 $|P_S| \geq 1$ かつ $\{t_i, t_{i+3}\} \subset P$ を満たす状態 P は, $(|Q_S|, |Q_T|) = (|P_S|-1, |P_T|)$, $(|P_S|, |P_T|-1)$, $(|P_S|, |P_T|-2)$ のいずれかを満たす状態 Q から到達可能である.

補題5 $|P_S| \geq 1$ かつ $\{t_i, t_{i+3}\} \subset P$ でない状態 P は, $(|Q_S|, |Q_T|) = (|P_S|-1, |P_T|)$ または $(|P_S|-1, |P_T|+1)$ を満たす状態 Q から到達可能である.

補題6 DFA M_1' において, 到達可能な状態は全て互いに同値でない.

補題1において, $3k+1$ 個の状態が到達不可能であることが示され, 補題2〜補題5と合わせ, その他の状態が全て到達可能であることが帰納的に示される. 補題6によって, この DFA が最小であることが示される.

4.3 補題1の証明

1. 空集合 ϕ については, 初期状態が一つあり, M_1 の全ての状態において, 0 および 1 による遷移が定義されているため, 明らかに到達不可能である.

2. $|P_T| = 1$, つまり $[t_i]$ と表される各状態は, 初期状態 $[t_0]$ から 0 を何度か読むことにより, つまり 0 遷移により到達可能である.

3. $|P_T| = 2$ を満たす状態 P については, $[t_i, t_{i+1}]$ という隣り合う (NFA の) 状態の組と $[t_i, t_{i+2}]$ という一つ飛ばしの状態の組を除き, 到達可能である.

片方の状態を t_0 に固定した $[t_0, t_i]$ の到達可能性が分かれば, その他の状態 $[t_i, t_{i+l}]$ の到達可能性も 0 遷移により分かる.

- $i = 4, 5, \cdots, k-4$ のとき,

$$[t_0] \xrightarrow{0^i} [t_i] \xrightarrow{1} [t_0, t_i]$$

- $i = 3, k-3$ のとき,

$$[t_0] \xrightarrow{0^3} [t_3] \xrightarrow{1} [s_0] \xrightarrow{0} [s_1] \xrightarrow{1} [t_0, t_3] \xrightarrow{0^{k-3}} [t_0, t_{k-3}]$$

以上から, $3 \leq i \leq k-3$ の i に対して, $[t_0, t_i]$ が到達可能であることが示された.

次に, 残った $[t_i, t_{i+1}]$, $[t_i, t_{i+2}]$ が到達不可能であることを述べる. まず, T 状態だけで遷移で到達可能であると仮定する. そのとき, これらの状態に到達するためには, $|P_T| = 1$, つまり $[t_i]$ と表される状態から状態数を増やす必要がある. T 状態のうち, 状態数が増える遷移を持っているのは, $[t_i] \xrightarrow{1} [t_i]$, $[t_i] \xrightarrow{1} [t_0]$ をともに持っているものだけである. $i = 1, 2, k-1, k-2$ を満たすいずれの i もそれに該当しないため, T 状態のみでの遷移は不可能であることが分かる.

そこで, S 状態を含む状態 K から遷移して到達することを考える. S 状態は 0 遷移で必ず S 状態に移るため, T 状態のみになるためには 1 で遷移する必要がある. S 状態から 1 を読んで T 状態に移る遷移は $[s_1] \xrightarrow{1} [t_0, t_3]$ のみであり, これは $[t_i, t_{i+1}]$, $[t_i, t_{i+2}]$ のいずれにも該当しない.

以上から, S 状態を含まない状態, および含む状態いずれからも到達することができないので, $[t_i, t_{i+1}]$, $[t_i, t_{i+2}]$ は到達不可能であることが示された.

4. $|P_T| = 3$ を満たす P については, $[t_i, t_{i+1}, t_{i+2}]$ という隣り合う 3 つの組を除いて, 全て到達可能である. まず, $|P_T| = 2$ の場合と同様に, 一つの状態を t_0 に固定して $[t_0, t_j, t_{j+r}]$ と表す. 3 で示したのと同様に, 0 遷移によって三つの状態の距離関係は変化しないまま T の環上で遷移するため, $j > 0$, $j + r < k$, $r > 0$ として一般性を失わない.

以下, T 状態のみの状態 P に関して, その中に含まれる t_i のうち, 0 遷移によって最も離れているもの同士の距離を P の "幅" と呼ぶことにする. 例えば, $[t_0, t_1]$ は幅 1 であり, $[t_0, t_1, t_2]$ は幅 2 である. 幅の大きさによって $[t_0, t_j, t_{j+r}]$ を分類し, それぞれについて証明を行う.

(a) 幅が 4 以下の状態

状態の種類は, 四点から異なる二点を選ぶ組合わせの数に等しく, 次の六通りに分類される: $[t_0, t_1, t_2]$, $[t_0, t_1, t_3]$, $[t_0, t_1, t_4]$, $[t_0, t_2, t_3]$, $[t_0, t_2, t_4]$, $[t_0, t_3, t_4]$. 以上のうち, 仮定の通り $\{t_i, t_{i+3}\}$ を含まないものは, $[t_0, t_1, t_2]$, $[t_0, t_2, t_4]$ の二つである.

$[t_0, t_1, t_2]$ は幅 2 の状態であるが, 1 遷移によって幅が 3 未満の T の状態に移るものは, T 上にも S 上にも存在しない. よって, $[t_0, t_1, t_2]$ は到達不可能である.

一方, $[t_0, t_2, t_4]$ は,

$$[t_0] \xrightarrow{0^4} [t_4] \xrightarrow{1} [t_0, t_2, t_4]$$

という遷移により到達可能である.

(b) 幅が 5 以上の状態

$5 \leq j \leq k-5$ であれば, 幅が 5 以上となる. このとき, $[t_j, t_{j+r}]$ が到達可能ならば, 直ちに $[t_j, t_{j+r}] \xrightarrow{1} [t_0, t_j, t_{j+r}]$ となり, 到達可能であることが分かる (但し, $5 \leq j \leq k-4$ を満たす t_j が $t_j \xrightarrow{1} t_0$ なる遷移を持つことを利用). 先ほど示した補題 1 の 3 よ

り，$r \geq 3$ ならば，$[t_j, t_{j+r}]$ は到達可能である．$r \leq 2$ のときは，$[t_j, t_{j+r}]$ は到達不可能であるが，$[t_0, t_r, t_{k-j}] \xrightarrow{0^j} [t_0, t_j, t_{j+r}]$，$[t_0, t_{k-j-r}, t_{k-r}] \xrightarrow{0^{j+r}} [t_0, t_j, t_{j+r}]$ であることから，もし t_{k-j}, t_{k-j-r} いずれかが t_5〜t_{k-4} に含まれていれば，$[t_{k-j}, t_r]$ または $[t_{k-j-r}, t_{k-j}]$ から，$r \geq 3$ と同様の遷移を経て $[t_0, t_j, t_{j+r}]$ に到達可能である．このことは，以下のように示すことができる．$5 \leq j \leq k-5$ より，$5 \leq k-j \leq k-5$ であるから，$j = 5$ の場合を除いて，t_{k-j} が t_5〜t_{k-4} に含まれる．そのときは 1 遷移によって $[t_0, t_j, t_{j+r}]$ に到達可能である．$j = 5$ のとき，$k = 9$ であると $[t_0, t_5, t_{5+r}]$ は幅が 4 になってしまうため，k が奇数であることから，$k \geq 11$ としてよい．すると，$4 \leq 6-r \leq k-j-r \leq k-5-r \leq k-6$ となり，$k-j-r = 4$ の場合を除き，$t_{k-j-r}, t_{k-j}]$ が t_5〜t_{k-4} に含まれることが分かる．よって，これらも 1 遷移によって到達可能である．最後に，$k-j-r=4$ のときは，条件 $j = 5$, $k \geq 11$, $r \leq 2$ を満たすのは $k = 11$, $r = 2$ のときのみであり，$[t_0, t_5, t_7]$ が該当する状態である．この状態は，$[t_2, t_6] \xrightarrow{1} [t_0, t_2, t_6] \xrightarrow{0^{13}} [t_0, t_4, t_9] \xrightarrow{0^7} [t_0, t_5, t_7]$ により，到達可能である．

以上より，補題 1 が示された． □

4.4 補題 2 の証明

$\{t_i, t_{i+3}\} \subset P$ より P は，$\{t_0, t_3\}$ を含み S 状態を持たない状態 P' から 0 遷移で到達可能である．そのような P' を一つ定める．ここで，$Q = (P' \backslash \{t_0, t_3\}) \cup \{s_1\} \xrightarrow{1} P'$, $t_4 \xrightarrow{1} [t_0, t_2, t_4]$ という遷移が成り立つため，P' が t_4 を含まないか，または t_2 と t_4 を含むならば，Q から P に到達可能である．

次に，P' が t_4 を含み，t_2 を含まないとき，$P = \{t_0, t_3, t_4\} \cup P'$ と表す (但し，$t_2 \notin P'$)．

このとき，ある i が存在して $t_{4i} \in P'$ かつ $t_{4(i+1)} \notin P'$ である場合とそのような i が存在しない場合が考えられるが，後者はありえない．なぜなら，もし，どのような自然数 i に対しても $t_{4i} \in P'$ であるとすると，k が奇数であることから，ある i が存在して，$t_{4i} = t_2$ となる．これは $t_2 \notin P'$ という仮定に反する．よって，以後 $t_{4i} \in P'$ かつ $t_{4(i+1)} \notin P'$ とする．さらに，$t_{4i+3} \in P'$ か否かによって場合分けする．

(a) $t_{4i+3} \in P'$ のとき，$P'' = \delta'^{-1}(P', 0^{4i})$，つまり，$P'$ の t_{4i} が t_0 に，t_{4i+3} が t_3 に移されるように P' に 0 遷移を施した状態を P'' とする．$\{t_0, t_3\} \subset P''$ かつ $t_4 \notin P''$ である．同様に，$R = \delta^{-1}(\{t_0, t_3, t_4\}, 0^{4i})$ とする．$t_{4(i+1)} \notin P'$ より，$t_4 \notin R$ である．すると $P = R \cup P'$ は $Q = R \cup (P'' \backslash \{t_0, t_3\}) \cup \{s_1\}$ から次のように遷移可能である．

$$Q \xrightarrow{1} R \cup P'' \xrightarrow{0^{4i}} P$$

$|Q_S| = 1$, $|Q_T| = |P_T| - 2$ である．

(b) $t_{4i+3} \notin P'$ のときも，(a) と同様，$P'' = \delta'^{-1}(P', 0^{4i})$ とする．すると，$t_0 \in P''$ かつ $t_3, t_4 \notin P''$ である．$R = \delta'^{-1}(\{t_0, t_3, t_4\}, 0^{4i})$ とする．$t_{4i+3}, t_{4(i+1)} \notin P'$ より，$t_3, t_4 \notin R$ である．すると $P = R \cup P'$ は $Q = R \cup (P'' \backslash \{t_0\})$ より，次のように遷移可能である．

$$Q \xrightarrow{1} R \cup P'' \xrightarrow{0^{4i}} P$$

なお，t_0 に対する遷移は，(a) と異なり，$\delta'^{-1}(t_0, 0^{4i})$ から行われる．この Q に関しては，$|Q_S| = 0$, $|Q_T| = |P_T| - 1$ である．

以上より，補題 2 が示された． □

4.5 補題 3 の証明

P は，$\{t_0\}$ を含み S の状態を持たない状態 P' から 0 遷移によって到達可能である．補題 2 と同様，まず P' が「t_4 を含み，t_2 を含まない状態でない (*)」場合に到達可能性を示す．

P' に t_5〜t_{k-4} のいずれかが含まれていれば，$Q = P' \backslash \{t_0\} \xrightarrow{1} P'$ という題意を満たす遷移が存在する．P' に t_5〜t_{k-4} のいずれも含まれていなければ，条件 (*) および $t_0 \in P'$, $|P'_T| \geq 4$, $\{t_i, t_{i+3}\}$ は P' に含まれない，という全ての条件を満たす状態は $[t_{k-2}, t_0, t_2, t_4]$ のみである．この状態は，

$$[t_0, t_2, t_4] \xrightarrow{0^2} [t_2, t_4, t_6] \xrightarrow{1} [t_0, t_2, t_4, t_6]$$

$$\xrightarrow{0^{k-2}} [t_{k-2}, t_0, t_2, t_4]$$

により到達可能である．

t_4 を含み t_2 を含まない状態 P' に関しては，$\{t_i, t_{i+3}\}$ が P に含まれないことより，補題2の(b)と同じ議論が適用でき，P が $|Q_S| = 0$, $|Q_T| = |P_T| - 1$ を満たす状態 Q から到達可能であることが示される．□

4.6 補題4の証明

$|P_S|$ の値により二通りに場合分けする．

1. $|P_S| \leq m-1$ のとき

$\{t_i, t_{i+3}\} \subset P$ である P は，$\{t_0, t_3, s_0\}$ を含み，s_1 を含まず，かつ $|P'_S| = |P_S|$ となる状態 P' から，0遷移によって到達可能である．但し，k と m が素でなければ，P' から P に到達できない場合があることに注意する．P' が「t_4 を含み t_2 を含まない状態」でない場合，

$$Q = (P' \backslash \{t_0, s_0\}) \cup \{s_1\} \xrightarrow{1} P'$$

により，Q より到達可能である．$|Q_S| = |P_S|$, $|Q_T| = |P_T| - 1$ である．

P' が「t_4 を含み t_2 を含まない状態」である場合，補題2と同様に，ある i が存在して，$t_{4i} \in P'$, $t_{4(i+1)} \notin P'$ とする．

(a) $t_{4i+3} \in P'$ のとき，k と m が互いに素であることを利用し，0遷移によって，P' を，t_{4i} は t_0 に移し，かつ，$s_0 \in P''$, $s_1 \notin P''$ であるような状態 P'' に移す（さらに $t_3 \in P''$, $t_4 \notin P''$ に注意する）．すると

$$Q = (P'' \backslash \{t_0, t_3, s_0\}) \cup \{s_1\} \xrightarrow{1} P''$$

により，P'' は（よって P も）Q より到達可能である．$|Q_S| = |P_S|$, $|Q_T| = |P_T| - 2$ である．

(b) $t_{4i+3} \notin P'$ のとき，k と m が互いに素であることを利用し，0遷移によって，P' を，t_{4i} は t_0 に移し，かつ，$s_0 \in P''$, $s_1 \notin P''$ であるような状態 P'' に移す（さらに $t_3, t_4 \notin P''$）．すると

$$Q = (P'' \backslash \{t_0, s_0\}) \cup \{t_3\} \xrightarrow{1} P''$$

により，P'' は（よって P も）Q より到達可能である．$|Q_S| = |P_S| - 1$, $|Q_T| = |P_T|$ である．

2. $|P_S| = m$ のとき

P は，$\{t_0, t_3\}$ および S を含む状態 P' から0遷移によって到達可能である．P' が「t_4 を含み t_2 を含まない状態」でない場合，

$$Q = P' \backslash \{t_0\} \xrightarrow{1} P'$$

となる．$|Q_S| = |P_S|$, $|Q_T| = |P_T| - 1$ である．

P' が「t_4 を含み t_2 を含まない状態」である場合，これまでと同様に，(a), (b) の場合分けを行う．

(a) $t_{4i}, t_{4i+3} \in P'$ かつ $t_{4(i+1)} \notin P'$ のとき，0遷移によって t_{4i} を t_0 に移すことで，P' から P'' を作る．$t_0, t_3 \in P''$, $t_4 \notin P''$, $P''_S = S$ である．このとき，

$$Q = (P'' \backslash \{t_0\}) \xrightarrow{1} P''$$

により，P'' は Q より到達可能である．$|Q_S| = |P_S| = m$, $|Q_T| = |P_T| - 1$ である．

(b) $t_{4i} \in P'$ かつ $t_{4i+3}, t_{4(i+1)} \notin P'$ のとき，0遷移によって t_{4i} を t_0 に移すことで，P' から P'' を作る．$t_0 \in P''$, $t_3, t_4 \notin P''$, $P''_S = S$ である．このとき，

$$Q = (P'' \backslash \{t_0, s_1\}) \cup \{t_3\} \xrightarrow{1} P''$$

により，P'' は Q より到達可能である．$|Q_S| = |P_S| - 1$, $|Q_T| = |P_T|$ である．

以上より，補題4が示された．□

4.7 補題5の証明

補題4と同様，$|P_S|$ の値により二通りに場合分けする．

1. $|P_S| \leq m-1$ のとき

$\{t_i, t_{i+3}\} \subset P$ であるため，P は，$\{t_0, s_0\}$ を含み $\{t_3, s_1\}$ を含まず，かつ $|P'_S| = |P_S|$ であるような P' から0遷移によって到達可能である．さらに，P' が「t_4 を含み，t_2 を含まない状態」でない場合，

$$Q = (P' \backslash \{s_0\}) \cup \{t_3\} \xrightarrow{1} P'$$

により Q より到達可能である．$|Q_S| = |P_S| - 1$, $|Q_T| = |P_T| + 1$ である．

P' が「t_4 を含み t_2 を含まない状態」である場合，$\{t_i, t_{i+3}\} \subset P'$ でないため，$t_{4i} \in P'$ かつ $t_{4i+3}, t_{4(i+1)} \notin P'$ の場合のみ考えればよい．

0遷移によって P' を次の条件を満たす状態 P'' に移す．t_{4i} は t_0 に移り，$t_0, s_0 \in P''$,

$t_3, t_4, s_1 \notin P''$. これは 4.6 節, 1(b) と同じケースであるので, P'' は $|Q_S| = |P_S| - 1$, $|Q_T| = |P_T|$ を満たす状態 Q から到達可能である.

2. $|P_S| = m$ のとき

P は, $\{t_0\}$ と全ての S 状態を含み, $\{t_3\}$ を含まず, かつ $|P'_S| = |P_S|$ であるような P' から 0 遷移によって到達可能である. さらに, P' が「t_4 を含み, t_2 を含まない状態」でない場合,

$$Q = (P' \backslash \{s_1\}) \cup \{t_3\} \xrightarrow{1} P'$$

により Q から到達可能である. $|Q_S| = |P_S| - 1$, $|Q_T| = |P_T| + 1$ である.

P' が「t_4 を含み t_2 を含まない状態」である場合, 1 と同様, $t_{4i} \in P'$ かつ $t_{4i+3}, t_{4(i+1)} \notin P'$ の場合のみ考えればよい. 0 遷移によって t_{4i} は t_0 に移すことで, P' から P'' を作る. すると, $t_0 \in P''$, $t_3, t_4 \notin P''$, $P''_S = S$ である. これは, 今考えている 2 の最初の「t_4 を含み, t_2 を含まない状態」でない場合と同じ条件なので, $|Q_S| = |P_S| - 1$, $|Q_T| = |P_T| + 1$ である状態 Q から到達可能である.

以上より, 補題 5 が示された. □

4.8 補題 6 の証明

ここでは, NFA M_1 と等価な DFA の到達可能な状態が全て互いに同値でないことを示す.

今, DFA の二つの異なる状態を X, Y とする. 仮に $X_T \neq Y_T$ であるとすると, $t_0 \in \delta'(X_T, 0^j)$ かつ $t_0 \notin \delta'(Y_T, 0^j)$ を満たす j が存在しなければならない. よって, このとき X と Y は同値でない. また, $X_T = Y_T$, $X_S \neq Y_S$ であるとすると, $s_1 \in \delta'(X_S, 0^j)$ かつ $s_1 \notin \delta'(Y_S, 0^j)$ を満たす j が存在しなければならない. ここから 1 遷移を施すと, $t_0 \in \delta'(X, 1)$ かつ $t_0 \notin \delta'(Y, 1)$ となるので, やはり X と Y は同値でない.

以上から, $X = Y$ であるときに限り X と Y は同値となり, 到達可能な全ての状態の非同値性が示された. □

4.9 $\alpha = 3k + 2, 3k + 3$ の場合

ここまでの証明で, $\alpha = 3k+1$ に対して, 等価な最小 DFA が $2^n - (3k+1)$ という状態数を持つような n 状態 NFA が構築可能であることを示したが, $\alpha \leq 3n-3$ である任意の α に対して, 同様の主張を示すには, $\alpha = 3k+2, 3k+3$ に対しても, 等価最小の DFA が $2^n - \alpha$ 状態を持つ NFA を構築する必要がある. $\alpha = 3k+2$ に対する NFA M_2 は, M_1 に遷移 $s_0 \xrightarrow{1} t_0$ を加えたものであり, $\alpha = 3k+3$ に対する NFA M_3 は, M_2 にさらに遷移 $t_{k-2} \xrightarrow{1} t_0$ を加えたものである.

M_2, M_3 に対する等価な最小 DFA の状態数を示すのが以下の二補題である.

補題 7 NFA M_2 を等価変換した DFA M'_2 においては, 定理 1 の DFA M'_1 に対して, 新たに $|P_T| = 0$, $|P_S| = m$ を満たす状態 P が到達不可能となり, 他の状態の到達可能性は M'_1 と一致する.

補題 8 NFA M_3 を等価変換した DFA M'_3 においては, DFA M'_2 に対して, 新たに $|P_T| = k$, $|P_S| = 0$ を満たす状態 P が到達不可能となり, 他の状態の到達可能性は M'_2 と一致する.

補題 7 の略証: 補題 5 より, $|P_T| = 0$, $|P_S| = m$ を満たす P には, t_3 を含み, s_1 を含まない状態 Q から 1 遷移により移す必要がある. 1 遷移によって $|P_S| = m$ となるためには, 1 遷移で s_1 に移れる状態が Q に含まれていなければならないが, それを唯一満たす s_0 からの 1 遷移を用いる $|P_T| = 1$ となってしまうため, $|P_T| = 0$, $|P_S| = m$ を満たす状態には到達できない.

他の状態の到達可能性については, s_0 からの 1 遷移を利用する証明が補題 4 の (2) にあるが, $s_0, s_1 \in Q$ であり, $s_1 \xrightarrow{1} t_0$ が既に存在するため, $s_0 \xrightarrow{1} t_0$ が付加されても, その他の状態の到達可能性には影響を与えない. よって, 補題 7 が示された. □

補題 8 の略証: $|P_T| = k$ を満たすためには, $|Q_T| < k$ である状態 Q から 1 遷移を行う必要がある. 仮に $t_{k-2} \in Q$ とすると, 1 遷移によって $|P_S| = 1$ となり題意を満たすことができない. よって, $t_{k-2} \notin Q$ である. しかし, 1 遷移で t_{k-2} に到達できるのは t_{k-2} 自身のみであるため, $|P_T| = k$, $|P_S| = 0$ を満たす状態 P に到達することは不可能である.

他の状態の到達可能性については，t_{k-2} を含む状態について，1遷移のみでは到達できない場合が生じる (s_0 への1遷移が存在するため)．その場合は，t_{k-2} を含まない状態 P'' にまず遷移し，その後0遷移により到達する．□

証明詳細はフルバージョンの論文に譲る．

注意：定理1の k が奇数という条件は，補題8で新たに付加された遷移 $t_{k-2} \xrightarrow{1} s_0$ から課される．すなわち，k が偶数であるときは，M_3 に関して，次の二状態 $[t_0, t_2, \cdots, t_{k-4}, t_{k-2}]$, $[t_1, t_3, \cdots, t_{k-3}, t_{k-1}]$ も到達不可能となる．その理由は以下の通りである．

$[t_0, t_2, \cdots, t_{k-4}, t_{k-2}]$ について考えると，この状態は，$\{t_i, t_{i+3}\}$ を持たないため，S 状態を含む P からは到達できない．よって，T 状態のみの状態から1遷移によって到達する必要がある．それには，$[t_2, t_4, \cdots, t_{k-4}, t_{k-2}]$, $[t_0, t_4, \cdots, t_{k-4}, t_{k-2}]$, $[t_4, \cdots, t_{k-4}, t_{k-2}]$ のいずれかの状態から1遷移によって移る必要がある．これら全てに t_{k-2} が含まれているため，1遷移を行うと s_0 に遷移してしまう．よって $[t_0, t_2, \cdots, t_{k-4}, t_{k-2}]$ は到達不可能である．また，$[t_1, t_3, \cdots, t_{k-3}, t_{k-1}]$ は $[t_0, t_2, \cdots, t_{k-4}, t_{k-2}]$ から0遷移で到達可能なので到達不可能である．

5 定理2の証明の概略

$\alpha = 3k+1, +2, +3$ の場合と同様にして，$\alpha = 4k+1, +2, +3$ に対するNFA M_4, M_5, M_6 を構成して証明を行う ($\alpha = 4k+4$ に対応するNFAがないことに注意する)．M_4, M_5, M_6 は，それぞれ M_1, M_2, M_3 から遷移 $t_4 \xrightarrow{1} t_2$ を除いたものである (但し，M_6 のみ，さらに $t_{k-2} \xrightarrow{1} s_0$ を $t_{k-1} \xrightarrow{1} s_0$ に置き換える)．すると，まず $[t_i, t_{i+2}, t_{i+4}]$ が到達不可能となる．さらに，1遷移によって自分自身と t_0 へと移る状態が定理1で t_5〜t_{k-4} であったものが t_4〜t_{k-4} となる．また，$t_4 \xrightarrow{1} t_2$ が存在しないため，$t_{k-2} \xrightarrow{1} s_0$ を $t_{k-1} \xrightarrow{1} s_0$ に変更することにより，P'' の条件を「t_0 を含み t_{k-1} を含まない」というものに変更できる．以上より，定理1と同様の証明が可能となる．

定理1の場合と同様に，α と k の式を k について解くと，$k = \lfloor (\alpha - 1)/4 \rfloor$ が得られる．

謝辞

本研究の一部は，文部科学省科学研究費補助金特定領域研究「新世代の計算限界 – その解明と打破 –」(課題番号16092215) の助成を受け行われました．この場を借りてお礼申し上げます．

参考文献

[1] M. Chrobak, Finite automata and unary languages, *Theoret. Comput. Sci.*, Vol. 47, pp. 149–158 (1986).

[2] V. Geffert, Magic numbers in the state hierarchy of finite automata, *Inform. and Comput.*, Vol. 205, No. 11, pp. 1652–1670 (2007).

[3] K. Iwama, Y. Kambayashi, K. Takaki, Tight bounds on the number of states of DFAs that are equivalent to n-state NFAs, *Theoret. Comput. Sci.*, Vol. 237, pp. 485–494 (2000).

[4] K. Iwama, A. Matsuura, M. Paterson, A family of NFAs which need $2^n - \alpha$ deterministic states, *Theoret. Comput. Sci.*, Vol 301, pp. 451–462 (2003).

[5] J. Jirásek, G. Jirásková, A. Szabari, Deterministic blow-ups of minimal nondeterministic finite automata over a fixed alphabet, *Proc. of DLT 2007*, pp. 254–265 (2007).

[6] G. Jirásková, Deterministic blow-ups of minimal NFA's, *Rairo Theoret. Inform. and Appl.*, Vol. 40, pp. 485–499 (2006).

[7] M. Rabin, D. Scott, Finite automata and their decision problems, *IBM Res. Develop.*, Vol. 3, pp. 114–129 (1959).

[8] L. Van Zijl, Magic numbers for symmetric difference NFAs, *Intern. J. of Found. on Comput. Sci.*, Vol. 16, No. 5, pp. 1027–1038 (2005).