数理解析研究所講究録
第 876 巻 1994 年 69-82

高々 n 個の状態数をもつ有限オートマトンの Vapnik-Chervonenkis 次元について
On the Vapnik-Chervonenkis dimensions of finite automata with n states

石上 嘉康 (Yoshiyasu Ishigami)*
谷 聖一 (Sei'ichi Tani)

Abstract

本論文では、状態数が高々 n である有限オートマトンによって受理される言語族の Vapnik-Chervonenkis 次元を考察する。$\mathrm{DFA}_{k,n}$ でアルファベットサイズが k で状態数が高々 n の決定性有限オートマトンによって受理される言語族を表し、$\mathrm{NFA}_{k,n}$ でアルファベットサイズが k で状態数が高々 n の非決定性有限オートマトンによって受理される言語族を表すものとする。本論文では、固定された任意の正整数 k に対して、(1) $\mathrm{DFA}_{1,n}$ の Vapnik-Chervonenkis 次元は $(1+o(1))n$、(2) $k \geq 2$ のとき、$\mathrm{DFA}_{k,n}$ の Vapnik-Chervonenkis 次元は $(k-1+o(1))n\log_2 n$、(3) $k \geq 2$ のとき、$\mathrm{NFA}_{k,n}$ の Vapnik-Chervonenkis 次元は $\Theta(n^2)$ が示される。

1 はじめに

本論文では、決定性有限オートマトン、非決定性有限オートマトンの Vapnik-Chervonenkis 次元 [15] について考察する。Vapnik-Chervonenkis 次元は PAC-学習 [13, 14] における学習に必要な例示数との関連 [6, 5] や質問を用いた EXACT-学習 [1, 2, 3] における学習に必要な質問数との関連 [11, 12] などが知られており、(言語族を含む) 様々な概念のクラスの Vapnik-Chervonenkis 次元を決定することは、計算論的学習理論への応用だけを考えても重要な問題である。

$\mathrm{DFA}_{k,n}$ で、アルファベットサイズが k で状態数が高々 n の決定性有限オートマトンによって受理される言語族を表し、$\mathrm{NFA}_{k,n}$ で、アルファベットサイズが k で状態数が高々 n の非決定性有限オートマトンによって受理される言語族を表すものとする。Σ_2 をある 2-文字アルファベットとし、${\Sigma_2}^{=n}$ で Σ_2 上の長さ n の語すべてでかつそれらのみの集合を表すものとし、$f(n)$ で ${\Sigma_2}^{=n}$ の部分集合を受理する最小状態数決定性有限オートマトンのうち状態数が最大のものの状態数とする。J. M. Champarnaud と J. E.

*日本国政府の文部省科学研究費補助金 (学術振興会特別研究員) の援助に感謝します。

Pin[7] は $1 = \liminf_{n\to\infty} nf(n)/2^n \leq \limsup_{n\to\infty} nf(n)/2^n = 2$ であることを示した。この結果から VC-dim($\mathrm{DFA}_{2,n}$) $\geq 1/2 \cdot n\log_2 n$ が導かれる。T. Gaizer[10] はそれとは独立に VC-dim($\mathrm{DFA}_{k,n}$) $= \Theta(n\log_2 n)$ を示している。本論文では、VC-dim($\mathrm{DFA}_{k,n}$) の漸近的な振る舞いを決定する次の結果が示される:

(1) VC-dim($\mathrm{DFA}_{1,n}$) $= (1+o(1))n$、
(2) 固定された任意の 2 以上の整数 k に対して、
$$\text{VC-dim}(\mathrm{DFA}_{k,n}) = (k-1+o(1))n\log_2 n。$$

非決定性有限オートマトンに関しては、固定された任意の整数 $k \geq 2$ に対して、$(k-1)n^2 \leq$ VC-dim($\mathrm{NFA}_{k,n}$) $\leq kn^2$であることが示される。

Vapnik-Chervonenkis 次元と計算論的学習理論との関連について、様々なことが知られている [6, 5, 12]。Maass と Turán [12] は質問数の複雑さ (query complexity) — ある概念族から未知の概念を同値性質問と所属性質問を用いて EXACT-学習するのに必要な質問数と Vapnik-Chervonenkis 次元の関連について次のような結果を得ている:

> $\mathcal{C}$ を有限の全体集合上の概念としてする。$\mathcal{C}$ を同値性質問と所属性質問を用いて EXACT-学習するのに必要な質問数はΩ(VC-dim($\mathcal{C}$)) である。

本稿では、Maass と Turán の結果の拡張と $\mathrm{DFA}_{k,n}$ の Vapnik-Chervonenkis 次元の下界を用いて、決定性有限オートマトンを同値性質問と所属性質問を用いて EXACT-学習するのに必要な質問数は学習目標の最小状態数決定性有限オートマトンの状態数を n とすると $\Omega(n\log_2 n)$ であることが示される。

D. Angluin [1] は決定性有限オートマトンを同値性質問と所属性質問を用いて、$O(n)$ 回の同値性質問と $O(mn^2)$ 回の所属性質問で EXACT-学習するアルゴリズムを示した。ただし、n は学習目標の最小状態数決定性有限オートマトンの状態数で、m は学習中に与えられる反例の最大長とする。Balcázar ら [4] は、制限学習モデル (bounded learning model) [16] において所属性質問を指数回行うことなく同値性質問を $O(\log_2 n)$ 回に減らすことができないことを示した。しかしながら、決定性有限オートマトンを学習する際に $O(n)$ 回の同値性質問を行える場合、所属性質問を何回行わなければならないかについてはこれまで、知られていなかった。

2 準備

本論文では、形式言語理論で標準的な教科書 [8] に従った定義や表記を用いる。Σ をあるアルファベットとする。語 とは Σ の元の有限列のことであり、Σ^* で語全体を表す。言語 とは Σ^* の部分集合のことである。λ で空語を表し、 語 x の長さを $|x|$ で表す。$\Sigma^{\leq m}$ と $\Sigma^{=m}$ は、それぞれ、$\{x \in \Sigma^* : |x| \leq m\}$ と $\{x \in \Sigma^* : |x| = m\}$ のことを表すものとする。任意の Σ^* の語 x, y に対して、$x \cdot y$ で x と y の連接を表す。任意の記号 $a \in \Sigma$ と任意の語 $w \in \Sigma^*$ に対して、$T(w \cdot a) = w$ 、$last(w \cdot a) = a$ とする。

任意の語 $w \in \Sigma^*$ に対して、$Proj_{[i]}(w)$ で w の i 番目の記号を表し、$Pref_{[i]}(w)$ で w の長さ i の接頭語を表し、w^R で w の逆語を表すものとする。例えば、$Proj_{[2]}(10110) = 0$、

$Pref_{[3]}(10110) = 101$、$10110^R = 01101$。任意の集合 A と B に対して、$A \triangle B$ で A と B の対称差（$(A \cup B) - (A \cap B)$）を表すものとする。任意の集合 S に対して、$|S|$ で S の要素数を表すものとする。$\mathbf{N}$ で非負整数の集合を、$\mathbf{R}$ で実数の集合を表すものとする。任意の非負整数 n に対して、$[n] = \{i \in \mathbf{N} : 1 \leq i \leq n\}$ とする。また、本論文中では、任意の実数 $n > 0$ に対して、$\log n$ で $\log_2 n$ を表すものとする。

本論文中以下では、特に断わらない限り、k-文字アルファベットとして $\Sigma_k = [k]$ を用いる。

$f, g : \mathbf{N} \to \mathbf{R}$ を充分大きな n に対して、$f(n), g(n) > 0$ となる関数とする。ここで、オーダを表す記号 O, Ω, Θ, o を次のように定義する:

- ある定数 $c > 0$ と十分大きな n に対して、$f(n) \leq c \cdot g(n)$ なら、$f = O(g)$、
- ある定数 $c > 0$ と十分大きな n に対して、$f(n) \geq c \cdot g(n)$ なら、$f = \Omega(g)$ 、
- $f = O(g)$ かつ $f = \Omega(g)$ なら $f = \Theta(g)$、
- 任意の定数 $c > 0$ とそれに対する十分大きな n に対して、$f(n) \geq c \cdot g(n)$ なら、$f = o(g)$。

有限オートマトン

決定性有限オートマトンとは次のような 5 個組 $(Q, \Sigma, \delta, q_0, F)$ である。ただし、Q は状態の有限集合、Σ は有限のアルファベット、q_0 は Q の元で初期状態、$F \subseteq Q$ は最終状態の集合、δ は $Q \times \Sigma$ から Q への関数とする。$\mathrm{dfas}_{k,n}$ でアルファベットサイズが k で、状態数が高々 n の決定性有限オートマトンの集合を表すものとする。DFA_k でアルファベットサイズが k の決定性有限オートマトンによって受理される言語族を表わし、$\mathrm{DFA}_{k,n}$ でアルファベットサイズが k で、状態数が高々 n の決定性有限オートマトンによって受理される言語族を表すものとする。

非決定性有限オートマトンとは次のような 5 個組 $(Q, \Sigma, \delta, q_0, F)$ である。ただし、Q は状態の有限集合、Σ は有限のアルファベット、q_0 は Q の元で初期状態、$F \subseteq Q$ は最終状態の集合、δ は $Q \times \Sigma$ から 2^Q への関数とする。$\mathrm{nfas}_{k,n}$ でアルファベットサイズが k で、状態数が高々 n の非決定性有限オートマトンの集合を表すもとする。NFA_k でアルファベットサイズが k の非決定性有限オートマトンによって受理される言語族を表わし、$\mathrm{NFA}_{k,n}$ でアルファベットサイズが k で、状態数が高々 n の非決定性有限オートマトンによって受理される言語族を表すものとする。任意の有限オートマトン FA に対して、$L(\mathrm{FA})$ で FA によって受理される言語を表すものとする。

Vapnik-Chervonenkis 次元

$\mathcal{F}$ を 全体集合 X の部分集合の族とする。

定義 2.1 集合 $S \subseteq X$ が $\mathcal{F}$ により 細分される とは、任意の S の部分集合 S' に対して、$S' = S \cap F$ となる $\mathcal{F}$ の元 F が存在するときかつそのときに限る。

言い換えると、S が $\mathcal{F}$ によって細分されるとは、$\{S \cap F : F \in \mathcal{F}\}$ がベキ集合 2^S になるときかつそのときに限る。

定義 2.2 集合族 $\mathcal{F}$ の Vapnik-Chervonenkis 次元 とは、$\mathcal{F}$ によって細分される集合の最大サイズとする。(そのような有限の値が存在しない場合は、Vapnik-Chervonenkis 次元を無限とする。)

$\mathcal{F}$ の Vapnik-Chervonenkis 次元を VC-dim($\mathcal{F}$) と表すこともある。定義より簡単に、次の事実が得られる。

事実 2.1 $\mathcal{F}$ を要素数有限の集合族とする。VC-dim($\mathcal{F}$) $\leq \log |\mathcal{F}|$。

証明 $\mathcal{F}$ は要素数が VC-dim($\mathcal{F}$) の集合 S を細分する。任意の $s \in S$ に対して、$s = F \cap S$ となる $F \in \mathcal{F}$ が存在するので、$|\mathcal{F}| \geq 2^{|S|} = 2^{\text{VC-dim}(\mathcal{F})}$。よって、VC-dim($\mathcal{F}$) $\leq \log |\mathcal{F}|$。 □

学習可能性

本稿では D. Angluin[1] が導入した質問を用いた EXACT-学習モデルにおける質問の複雑さを扱う。まず、学習可能性を定義し、それから、質問の複雑さを定義する。このモデルでは、学習者の目標はある全体集合 X 上の固定された概念族 $\mathcal{C} \subseteq 2^X$ から未知の目標概念を、定められた種類の質問を行いながら見つけることである。本稿では、次の3つの質問を考える。学習者は未知の目標概念 $C_T \in \mathcal{C}$ を学習しようとしているとする。

1. 所属性質問 *(Membership query)* (Mem): 入力は語 $x \in \Sigma^*$、出力は $x \in C_T$ なら *yes* そうでなければ *no*;
2. 同値性質問 *(Equivalence query)* (Equ): 入力は仮説 $h \in \mathcal{C}$、出力は $h = C_T$ なら *yes* そうでなければ *no*。*no* の場合には、$h \triangle C_T$ の任意の元が反例 として与えられる;
3. 任意同値性質問 *(Arbitrary equivalence query)* (Arb): 入力は仮説 $h \in 2^X$、出力は $H = C_T$ なら *yes* そうでなければ *no*。ただし、*no* の場合には、$\Phi(h) \triangle C_T$ の任意の元が反例 として与えられる。入力となる仮説は $\mathcal{C}$ の元でなくても良いことに注意。

定義 2.3 $\mathcal{Q}$ を質問の集合とする。アルゴリズム A が $\mathcal{Q}$ を用いて $\mathcal{C}$ を学習するアルゴリズムであるとは、任意の目標概念 $t \in \mathcal{C}$に対して、A が t に関する $\mathcal{Q}$ に含まれる質問に正しく答える神託が与えられて実行されたならば、A は停止して $h = t$ となる仮説 h を出力するときかつそのときに限る。

本稿で扱う質問集合は {Equ,Mem} と {Arb,Mem} である。次に、"質問の複雑さ" を定義する。直観的には、質問の複雑さとは学習アルゴリズムが学習中に行わなければならない最悪の場合の質問数である。正確には、任意の概念族 $\mathcal{C}$、任意の質問集合 $\mathcal{Q}$、$\mathcal{Q}$ を用いて $\mathcal{C}$ を学習する任意のアルゴリズム A に対して、質問の複雑さ $\#query_{(A,\mathcal{Q})}$ は次

のように定義される:
任意の目標概念 $t \in C$ に対して、

$$\#query_{\langle A,Q\rangle}(t) = \max\{\mathrm{i} \in \mathbf{N}: \ A \text{ が停止して } h = t \text{ となる仮説 } h \text{ を出力するまでに } i \text{ 回の質問をしなくてはならない反例の選択が存在する。}\}$$

3 $\mathrm{DFA}_{k,n}$ の Vapnik-Chervonenkis 次元

この節では、$\mathrm{DFA}_{k,n}$ の Vapnik-Chervonenkis 次元について考察する。まず、最初に $k = 1$ の場合を考え、後から $k \geq 2$ の場合を考える。

集合 S を $\{1^i : i = 0, 1, \ldots, n-1\}$ とする。$|S| = n$ となり、S が $\mathrm{DFA}_{1,n}$ により細分されるのは明らかなので、下界 $\text{VC-dim}(\mathrm{DFA}_{1,n}) \geq n$ が得られる。事実 2.1より、$|\mathrm{DFA}_{1,n}|$ の上界を評価することにより、$\text{VC-dim}(\mathrm{DFA}_{1,n})$ の上界を得ることができる。

定理 3.1

$$|\mathrm{DFA}_{1,n}| \leq 2^n \cdot \frac{n^2+n}{2}$$

証明　この場合には、アルファベットの種類が 1 種類なので、ある語がある $\mathrm{DFA}_{1,n}$ の言語に含まれるかどうかは、その語の長さだけで決まる。ここで、少し、言葉の定義をする。f が集合 A の n-着色であるとは、f が A から $[n]$ への関数のときをいう。f が集合 A の n-分割であるとは、f が A から $[n]$ への関数のときをいう。ただし、A の分割として同じものは同一視する。たとえば、

$$B_1(x) = \begin{cases} 1 & \text{if } x = 0 \\ 2 & \text{otherwise} \end{cases}, \qquad B_2(x) = \begin{cases} 2 & \text{if } x = 0 \\ 1 & \text{otherwise} \end{cases}$$

としたとき、$B_1(n)$ と $B_2(n)$ は 2-分割として同一視する。$\{1,2\}$ の 2-着色は全部で 4 通りあるが、$\{1,2\}$ の 2-分割は全部で 2 通りである。$\mathrm{DFA}_{1,n}$ に含まれる任意の言語 L に対して、$\mathbf{N}$ の 2-着色 A_L を次のように定義する:

$$A_L(x) = \begin{cases} 1 & \text{if } 1^x \in L \\ 0 & \text{otherwise} \end{cases}$$

$\mathrm{DFA}_{1,n}$ の言語と $\mathbf{N}$ の 2-着色 が 1 対 1 に対応しているのは、簡単にわかる。ここで、我々が知りたいのは、状態数が高々 n の 1-文字正則言語 の数 $|\mathrm{DFA}_{1,n}|$ なので、任意の $\mathbf{N}$ の 2-着色 $A : \mathbf{N} \to [2]$ と $x \in \mathbf{N}$ に対して、$\mathbf{N}$ の 2-着色 $[x\backslash A]$ を次のように定義する: 任意の $y \in \mathbf{N}$ に対して、$[x\backslash A](y) = A(x+y)$。

補題 1

$$|\mathrm{DFA}_{1,n}| = |\{\mathbf{N} \text{ の 2-着色 } A : |\{[x\backslash A] : x \in \mathbf{N}\}| \leq n\}|。$$

証明　$L \subseteq {\Sigma_k}^*$ をある言語とする。任意の語 $u, v \in {\Sigma_k}^*$ に対してすべての ${\Sigma_k}^*$ の語 w に対し、$u \cdot w \in L \Leftrightarrow v \cdot w \in L$ となるとき、$u \equiv_L v$ と定義する。$[w]_{\equiv_L}$ で $\{w' \in {\Sigma_k}^* : w' \equiv_L w\}$ を表すものとする。L が正則言語なら $|\{[w]_{\equiv_L} : w \in \mathbf{N}^k\}|$ は有限で、かつ、$|\{[w]_{\equiv_L} : w \in \mathbf{N}^k\}|$ は L を受理する最小状態数決定性有限オートマトンの状態数に等しいことは良く知られている事実である。(例えば、文献 [8] 3.4 節を見よ。)

A_L は先に定義した $\mathbf{N}$ の 2-着色とする。任意の語 $w \in \Sigma_k$ に対して、$1^x \cdot w \in L \Leftrightarrow 1^y \cdot w \in L$ なので、$x, y \in \mathbf{N}$ に対して、$1^x \equiv_L 1^y$ が成り立つならば $[x\backslash A_L] = [y\backslash A_L]$ が成り立つ。もし、$1^x \not\equiv_L 1^y$ ならば、$A(x+w) \neq A(y+w)$ となるような語 $w \in {\Sigma_1}^*$ が存在するので、$[x\backslash A] \neq [y\backslash A]$ となる。それ故、$|\mathrm{DFA}_{1,n}| = |\{A : |\{[x\backslash A] : x \in \mathbf{N}\}| \leq n\}|$ となる。　△

以下では、$|\{[x\backslash A] : x \in \mathbf{N}\}| \leq n\}|$ となるような $\mathbf{N}$ の 2-着色の数を評価する。$B : \mathbf{N} \to [n]$ を $\mathbf{N}$ の n-分割とする。$x \in \mathbf{N}$ と B に対して、$\mathbf{N}$ の n-分割 $[x\backslash B]$ を次のように定義する: 任意の $y \in \mathbf{N}$ に対して、$[x\backslash B](y) = B(x+y)$。$\mathbf{N}$ の 2-着色 A に対して、次のような条件 $(*)$ を考える。

$(*)$　任意の $x, y \in \mathbf{N}$ に対して、$B_A(x) = B_A(y)$ ならば $[x\backslash B_A] = [y\backslash B_A]$。

A を $|\{[x\backslash A] : x \in \mathbf{N}\}| \leq n$ となる $\mathbf{N}$ の 2-着色とし、$\mathbf{N}$ の n-分割 B_A を次のように定義する: 任意の $x, y \in \mathbf{N}$ に対して、$[x\backslash A] = [y\backslash A]$ のときかつそのときに限り、$B_A(x) = B_A(y)$ となる。このように決めた分割は条件 $(*)$ を満たしている。(図 1 を参照せよ。)

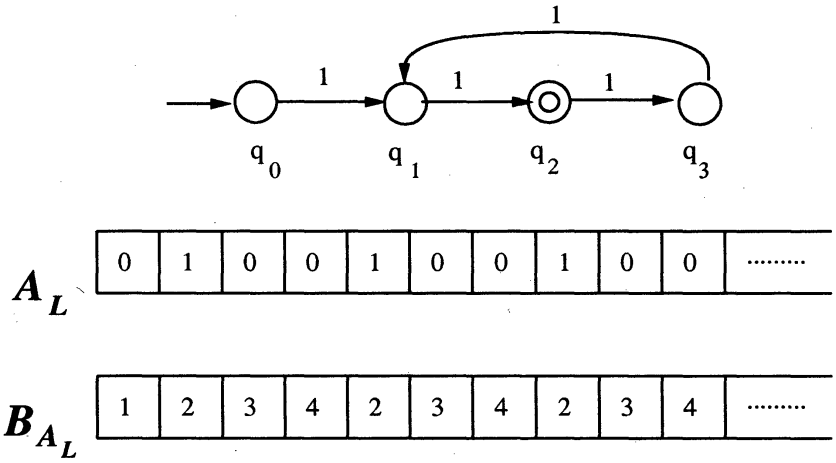

図 1: A_L と B_{A_L} の例 ($L = 1(111)^*$ の場合)

この事実から、$f(n)$ を

$$f(n) = |\{\ \mathbf{N} \text{ の } n\text{-分割 } B : B \text{ は条件 } (*) \text{ を満たす } \}|$$

と定義すると、

$$|\{\mathbf{N} \text{ の 2-着色 } A : |\{[x\backslash A] : x \in \mathbf{N}\}| \leq n\}| \leq 2^n \cdot f(n)$$

となり、補題 1 から次の補題 2 が得られる。

補題 2 $|\mathrm{DFA}_{1,n}| \leq 2^n \cdot f(n)$。

以下では、$f(n)$ を評価することになる。

補題 3

$$f(n) = \frac{n^2+n}{2}$$

証明　B を条件 $(*)$ を満たしている 1-次元無限配列の n-着色とする。$B(x_1) = B(x_2)$ かつ $x_1 < x_2$ ならば、任意の $x > x_2$に対して、$B(x) = B(x_1 + k)$。ただし、$k = (x - x_2) \bmod (x_2 - x_1)$。なぜなら、$B(x_1) = B(x_2)$ なので、任意の $j \in \mathbf{N}$ に対して、$B(x_1 + j) = B(x_2 + j)$ となり、$l = x_2 - x_1$ とおき、$m = (x - x_1 - k)/l$ とおくと、$B(x_1 + k) = B(x_2 + k) = B(x_1 + l + 1) = \cdots = B(x_1 + ml + k) = B(x)$ となるからである。つまり、n' $(\leq n)$ 個の分割が B に現れるなら、$B(0)$ から $B(n'-1)$ までに n' 個の分割が 1 つづつ全て現れているということである。条件 $(*)$ を満たすような分割を構成するため、$B(0)$ 、$B(1)$ 、...、と順番に分割を決めてゆき、$B(i-1)$ $(i \leq n-1)$ まで分割が決まっているとする。次に既に割り当てられている分割を割り当てると、先程の性質から、それ以降の分割は全て決まってしまう。まだ割り当てられていない分割を割り当てた場合は、さらに分割を続けることができる。よって、出現する分割の数 $n'(\leq n)$ を決めると、n' 個の分割は $B(0)$ から $B(n'-1)$ まで互いに違う n' 個の分割を割り当てたことになり、そのすぐ外側 $B(n')$ に n' 個の分割の中からどの分割を割り当てるか決めれば、条件 $(*)$ を満たす 1-次元無限配列の n-着色 が 1 つ決まる。よって、$f(n) = 1 + 2 + \cdots + n = (n^2+n)/2$。 △

補題 2 と補題 3 より直ちに $|\mathrm{DFA}_{1,n}| \leq 2^n \cdot (n^2+n)/2$ が導かれる。 □

系 3.2

$$\text{VC-dim}(\mathrm{DFA}_{1,n}) = (1 + o(1))n。$$

ここから、$k \geq 2$ の場合について考える。

定理 3.3 k を 2 以上の固定された整数とすると、

$$\text{VC-dim}(\mathrm{DFA}_{k,n}) = (k - 1 + o(1))n \log n。$$

証明

（上界）　事実2.1から、状態数が高々 n の決定性有限オートマトンに受理される言語の数を評価すれば良い。状態数が n の決定性有限オートマトンの数を考える。初期状態の取り方は n 通り、最終状態の取り方は 2^n 通り、遷移関数の取り方は入力が kn 通りで出力が

n 通りなので n^{kn} 通りとなり、状態数が n の決定性有限オートマトンの数は $n \cdot 2^n \cdot n^{kn}$ 通りとなる。この状態数が n の決定性有限オートマトンの数え方において、$\mathrm{DFA}_{k,n}$ に含まれる言語 L を受理する決定性オートマトンがどのくらい重複して数えられているか考える。ある正則言語 L を受理する最小状態数決定性有限オートマトンを M_L とする。M_L は最小状態数決定性オートマトンなので各状態は、$[w]_{\equiv_L}$ （$[w]_{\equiv_L}$ は6ページの補題1 の証明中に定義されている）の互いに違う元に対応している。つまり、L を受理する最小状態数決定性有限オートマトンの状態数が n のときは、M_L の状態を順列してできる決定性有限オートマトンは、この数え方においては互いに違う決定性有限オートマトンになっており、L を受理する決定性有限オートマトンは $n!$ 回数えられている。L を受理する最小状態数決定性有限オートマトンの状態数が $l(< n)$ の場合は、L を受理する最小状態数決定性有限オートマトンで l 個の状態が $q_1, q_2, \ldots, q_l$ とラベルづけされているもの M_L をもとに次のような決定性有限オートマトン $M_L' = ([k], Q, \delta, q_{init}, F)$ を考える:

- $Q = \{q_1, q_2, \ldots, q_n\}$ は状態の集合、
- $\delta : Q \times [k] \to Q$ は次のように定義された遷移関数:
  1. 任意の $i \in \{1, 2, \ldots, l\}$ と任意の $a \in [k]$ に対して、$\delta(q_i, a)$ は M_L と全く同様に定義する、
  2. 任意の $i \in \{l+1, l+2, \ldots, n-1\}$ と任意の $a \in [k]$ に対して、$\delta(q_i, a) = q_{i+1}$、
  3. 任意の $a \in [k]$ に対して、$\delta(q_n, a) = q_n$、
- q_{init} は初期状態で M_L の初期状態と同じものとする、
- F は最終状態の集合で M_L の初期状態と同じものとする。

すると、M_L' の状態を順列してできる $n!$ 個の決定性有限オートマトンすべてが、この数え方では別々に数えられているのが簡単にわかる。よって、任意の言語 $L \in \mathrm{DFA}_{k,n}$ に対して、L を受理する決定性オートマトンはこの数え方で少なくとも $n!$ 回重複して数えられていることになる。よって、状態数が n の決定性有限オートマトンに受理される言語の数は、高々 $2^n \cdot n^{kn}/(n-1)!$ となり、状態数が高々 n の決定性有限オートマトンに受理される言語の数も、高々 $2^n \cdot n^{kn}/(n-1)!$ となる。

$$
\begin{aligned}
\log\left(\frac{2^n \cdot n^{kn}}{(n-1)!}\right) &\sim \log\left(\frac{2^n \cdot n^{kn}}{\sqrt{2\pi(n-1)} \cdot (n-1)^{(n-1)}/e^n}\right) \\
&= n + kn\log n + n\log e - \log\sqrt{2\pi(n-1)} - (n-1)\log(n-1) \\
&\leq n + n\log e + (k-1)n\log n + n(\log n - \log(n-1)) + \log n \\
&= \left(k - 1 + \frac{1 + \log e + \log n - \log(n-1)}{\log n} + \frac{1}{n}\right) n\log n。
\end{aligned}
$$

よって、ある定数 C に対して、$\mathrm{VC\text{-}dim}(\mathrm{DFA}_{k,n}) \leq (k - 1 + C/\log n) n \log n$。

(下界) $N_1 = n - \lfloor n/\log n \rfloor$、$N_2 = \lfloor n/\log n \rfloor$ とおく。${\Sigma_k}^*(=[k]^*)$ の部分集合 W_1, W_2 を次のように定義する:

$$\begin{aligned} W_1 &= \left\{ w \in [k]^* : \textstyle\sum_{i=1}^{|w|} k^{|w|-i} \cdot Proj_{[i]}(w) < N_1 \right\}, \\ W_2 &= \left\{ w \in [k]^* : \exists a \in [k] \text{ such that } \textstyle\sum_{i=1}^{|w|} k^{|w|-i+1} \cdot Proj_{[i]}(w) + a \geq N_1 \right\}. \end{aligned}$$

$W = W_1 - W_2$ とする。

$$|W| = \frac{k-1}{k}\left(1 - \frac{1}{\log n}\right) n + o(n)$$

となり、W に含まれる任意の語 w_1, w_2 に対して、w_1 と w_2 は共通の接頭語を持たないことに注意せよ。${N_2}' = \lfloor \log(N_2+2) \rfloor - 1$ とおく。集合 S_k を $\{w \cdot a \cdot 1^i : w \in W, \quad a \in [k], \quad i = 0, 1, \ldots, {N_2}' - 1\}$ と定義する。$|S_k| = |W| \cdot k \cdot {N_2}' = (k - 1 + o(1)) n \log n$ となるのは容易に分かる。それ故、S_k が $\mathrm{DFA}_{k,n}$ によって細分されることを示せば良い。そのために、S_k の任意の部分集合 S' に対して、$L(M(S')) = S'$ となるような決定性有限オートマトン $M(S') \in \mathrm{dfas}_{k,n}$ を構成できることを示す。

任意の集合 $S' \subseteq S_k$、任意の語 $w \in W$、任意の記号 $a \in [k]$ に対して、${S'}_{w \cdot a}$ で $S' \cap \{w \cdot a \cdot 1^i : i = 0, 1, \ldots {N_2}' - 1\}$ を表すものとする。また、記号 $[{S'}_{w \cdot a}]$ で長さ ${N_2}'$ の次の条件を満たす語を表すものとする:

$$Proj_{[i]}([{S'}_{w \cdot a}]) = \begin{cases} 1 & \text{if } \ w \cdot a \cdot 1^{i-1} \in {S'}_w, \\ 0 & \text{otherwise.} \end{cases}$$

任意の $S' \subseteq S_k$ に対して、$M(S') = ([k], Q, \delta, q_{init}, F)$ を次のように定義する:

- $Q = A \cup B$ は状態の有限集合、ただし、$A = \{q_{\langle A,w \rangle} : w \in W_1\}$ かつ $B = \{q_{\langle B,w \rangle} : w \in \Sigma^{\leq {N_2}'} - \{\lambda\}\}$、
- $\delta : Q \times \Sigma \to Q$ は次のように定義された遷移関数:
  1. A、$w \in W_2$、$\sum_{k=1}^{|w|} k^{|w|-i+1} \cdot Proj_{[i]}(w) + a < N_1$ となる $a \in [k]$ に対して、
  $$\delta(q_{\langle A,w \rangle}, a) = q_{\langle A, w \cdot a \rangle},$$
  2. A、$w \in W$、$a \in [k]$ に対して、
  $$\delta(q_{\langle A,w \rangle}, a) = q_{\langle B, [{S'}_{w \cdot a}]^R \rangle},$$
  3. B、$w \in \Sigma^{\leq {N_2}'} - \Sigma^{\leq 1}$ に対して、
  $$\delta(q_{\langle B,w \rangle}, 1) = q_{\langle B, T(w) \rangle}$$
- $q_{init} = q_{\langle A,\lambda \rangle}$ は初期状態、
- $F = \{q_{\langle B,w \rangle} : q_{\langle B,w \rangle} \in B,\ last(w) = 1\}$ は最終状態の集合。

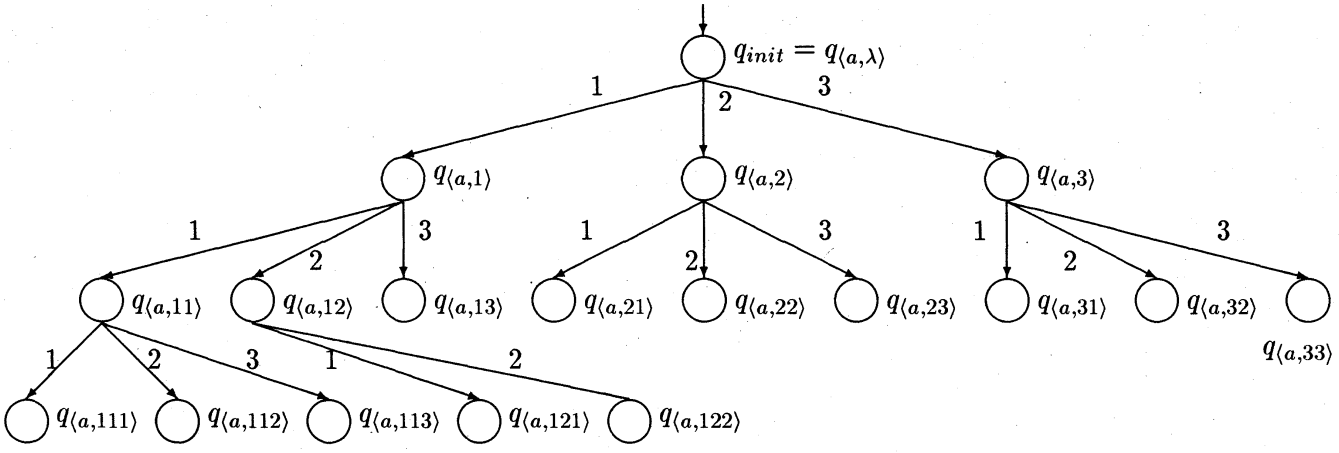

図 2: 遷移規則 (1) の $k = 3$ で $N_1 = 18$ のときの例

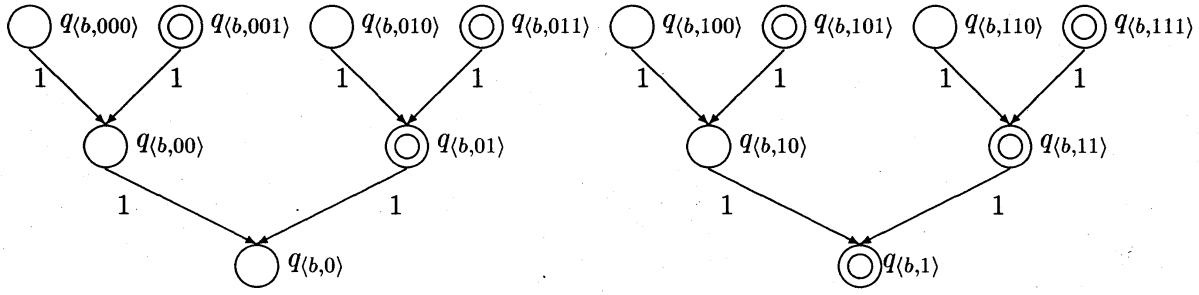

図 3: 遷移規則 (3) と最終状態の $N_2' = \lfloor N_2 + 2 \rfloor - 1 = 3$ のときの例

$A \times [k]$ から A への遷移規則 (1) と $B \times [k]$ から B への遷移規則 (3) とはすべての $S' \subseteq S$ に対応するオートマトン $M(S')$ の間で共通である。図 2 に遷移規則 (1) の例を、図 3 に遷移規則 (3) の例を示してある。

$A \times [k]$ から B への遷移規則 (2) は互いに違っている。

任意の $q \in Q$, $a \in [k]$, $w \in \Sigma^*$ に対して、$\delta(q, a \cdot w) = \delta(\delta(q, a), w)$ として δ の定義域を $Q \times [k]^*$ に拡張する。遷移規則 (1) の定義より、任意の W_1 の語 w に対して、$\delta(q_{init}, w) = q_{\langle A,w\rangle}$ となっている。最終状態は、B に含まれる状態のみで、初期状態は A に含まれ、A に含まれる状態から B に含まれる状態への遷移は、遷移規則 (2) でのみ与えられている。よって、$W \cdot [k]$ に含まれる語を接頭語として持たない語を $M(S')$ は受理しない。さらに、遷移規則 (3) より、$M(S')$ は S に含まれない語を受理しないのは明であろう。$w \cdot a \cdot 1^i$ を $w \in W$ で $w \cdot a \cdot 1^i \in S$ であるようなある語とする。δ の定義により、

$$\begin{aligned} \delta(q_{init}, w \cdot a \cdot 1^i) &= \delta(q_{\langle A,w\rangle}, a \cdot 1^i) \\ &= \delta(q_{\langle B,[S'_{w \cdot a}]^R\rangle}, 1^i) \\ &= q_{\langle B,Pref_{[N_2'-i]}([S'_{w \cdot a}]^R)\rangle}。 \end{aligned}$$

また、最終状態の集合 F の定義より、$w \cdot a \cdot 1^i$ が $M(S')$ に受理されるのは、$last(Pref_{[N_2{}'-i]}([S'_{w \cdot a}]^R)) = 1$ のときかつそのときに限る。$last(Pref_{[N_2{}'-i]}([S'_{w \cdot a}]^R)) = last(Pref_{[i]}([S'_{w \cdot a}])) = Proj_{[i]}([S'_{w \cdot a}])$ なので、$w \cdot a \cdot 1^i$ が $M(S')$ に受理されるのは、$w \cdot a \cdot 1^i \in S'_{w \cdot a}$ が成り立つときかつそのときに限る。つまり、$L(M(S')) = S'$ のときかつそのときに限る。それ故、S は $\mathrm{DFA}_{k,n}$ によって細分される。 □

4 $\mathrm{NFA}_{k,n}$ の Vapnik-Chervonenkis 次元

本節では、$\mathrm{NFA}_{k,n}$ の Vapnik-Chervonenkis 次元について考察する。

定理 4.1 n を十分大きな整数、k を固定された 2 以上の整数とする。

$$(k-1)n^2 \leq \text{VC-dim}(\mathrm{NFA}_{k,n}) \leq kn^2。$$

証明 (上界) 状態数が n の非決定性有限オートマトンの数を考える。初期状態の取り方は n 通り、最終状態の取り方は 2^n 通り、遷移関数の取り方は入力が kn 通りで出力が 2^n 通りなので 2^{kn^2} 通りとなり、状態数が n の非決定性有限オートマトンの数は $n \cdot 2^n \cdot 2^{kn^2}$ 通りとなる。決定性有限オートマトンの際（定理 3.3 の上界の証明）と同様に同じ言語を受理する非決定性有限オートマトンが少なくとも $n!$ 回重複して数えられているので、状態数が n の非決定性有限オートマトンに受理される言語の数は、高々 $2^n \cdot 2^{kn^2}/(n-1)!$ となり、状態数が高々 n の決定性有限オートマトンに受理される言語の数も、高々 $2^n \cdot 2^{kn^2}/(n-1)!$ となる。よって、$|\mathrm{NFA}_{k,n}| \leq 2^n \cdot 2^{kn^2}/(n-1)!$。よって、事実 2.1より、$\text{VC-dim}(\mathrm{NFA}_{k,n}) \leq \log(2^n \cdot 2^{kn^2}/(n-1)!)$ となり、$n \geq 6$ のとき、$\log(2^n \cdot 2^{kn^2}/(n-1)!) \leq kn^2$ となる。
(下界) アルファベット Σ_k 上の語の集合 S を $\{1^i a_l 1^j : i = 0, 1, \ldots, n-1, j = 0, 1, \ldots, n-1, a_l \in \Sigma_k - \{1\}\}$ とする。明らかに、$|S| = (k-1)n^2$。任意の集合 $S' \subseteq S$ に対して、S' に含まれる語はすべて受理し、$S - S'$ に含まれる語はすべて受理しない非決定性オートマトン $M_{S'} = (\Sigma_k, Q, q_{init}, \delta, F) \in \mathrm{nfas}_{k,n}$ を構成できることを示すことにより、S が $\mathrm{NFA}_{k,n}$ によって細分されることを示す。Q を $\{q_0, q_1, \ldots, q_{n-1}\}$ とし、$q_{init} = \{q_0\}$ とし、and $F = \{q_{n-1}\}$ とする。遷移関数 δ は以下のように定義する: (1) 任意の $i = 0, 1, \ldots, n-2$ に対して、$\delta(q_i, 1) = q_{i+1}$、(2) $1^i a_l 1^j \in S'$ なら$\delta(q_i, a_l) = q_{n-1-j}$。$M_{S'}$ は S' のすべての語を受理し、$S - S'$ のすべての語を受理しない。よって、S は $\mathrm{NFA}_{k,n}$ によって細分される。 □

5 有限オートマトンの EXACT-学習における質問数の複雑さ

本節では、有限オートマトンの EXACT-学習における質問数の複雑さについて考察する。Maass と Turán は有限の全体集合上の概念族を EXACT-学習するのに必要な質問数について次の結果を示している。

命題 5.1 ([12]) 質問の集合 $\mathcal{Q}$ を $\{\mathrm{Equ}, \mathrm{Mem}\}$ か $\{\mathrm{Arb}, \mathrm{Mem}\}$ のいずれかとする。任意の概念族 $\mathcal{C}$ と $\mathcal{Q}$ を用いて C を学習するアルゴリズム A に対して、

$$\max\{\#\mathrm{query}_{\langle \mathrm{A},\{\mathrm{Equ},\mathrm{Mem}\}\rangle}(\mathrm{t}) : \mathrm{t} \in \mathcal{C}\} = \Omega(\mathrm{VC\text{-}dim}(\mathcal{C})).$$

これらの結果は有限でない全体集合上の概念族 $\mathcal{C}$ の学習の場合にもなり立つので、次の補題が直ちに得られる。ρ は $\mathcal{C}$ から $\mathbf{N}$ へのサイズ関数とする。$\mathcal{C}_n$ を $\mathcal{C}_n = \{c \in \mathcal{C} : \rho(c) \leq n\}$ で定義する。

補題 5.2 任意の概念族 $\mathcal{C}$ に対して、$\{\mathrm{Equ}, \mathrm{Mem}\}$ を用いて A を学習するアルゴリズムで次の条件を満足するようなものは存在しない:
任意の目標概念 $t \in \mathcal{C}$ に対して、

$$\#query_{\langle A,\{\mathrm{Equ},\mathrm{Mem}\}\rangle}(t) = o\left(\mathrm{VC\text{-}dim}\left(\mathcal{C}_{\rho(\mathrm{t})}\right)\right)。$$

<u>証明</u>　任意の $t \in \mathcal{C}$ に対して、

$$\#query_{\langle A,\{\mathrm{Equ},\mathrm{Mem}\}\rangle}(t) = o(\mathrm{VC\text{-}dim}(\mathcal{C}_{\mathrm{n}}))$$

となるような学習アルゴリズム A が存在したとする。この A に $\{\mathrm{Arb},\mathrm{Mem}\}$ を用いて $\mathcal{C}_n$ を学習させる場合を考える。すると、任意の $t \in R_n$ に対して、A は t を $\#query_{\langle A,\{\mathrm{Arb},\mathrm{Mem}\}\rangle}(t) = o(\mathrm{VC\text{-}dim}(\mathcal{C}_{\mathrm{n}}))$ で学習できることになる。このことは、命題 5.1 に矛盾。 □

定理 3.3、定理 4.1 とこの補題から直ちに次の 2 つの系が得られる。

系 5.3 $\{\mathrm{Equ}, \mathrm{Mem}\}$ を用いて $\mathrm{DFA_k}$ を学習するアルゴリズムで次を満足するものは存在しない:
任意の目標決定性有限オートマトン t に対して、

$$\#query_{\langle A,\{\mathrm{Equ},\mathrm{Mem}\}\rangle}(t) \geq (k-1)n\log n。$$

ただし、n は t を受理する最小状態数決定性オートマトンの状態数。

系 5.4 $\{\mathrm{Equ}, \mathrm{Mem}\}$ を用いて $\mathrm{NFA_k}$ を学習するアルゴリズムで次を満足するものは存在しない:
任意の目標非決定性有限オートマトン t に対して、

$$\#query_{\langle A,\{\mathrm{Equ},\mathrm{Mem}\}\rangle}(t) \geq (k-1)n^2。$$

ただし、n は t を受理する最小状態数非決定性オートマトンの状態数。

参考文献

[1] Angluin, D.: Learning regular sets from queries and counterexamples, *Information and Computation*, Vol. 75, 1987, pp.87-106.

[2] Angluin, D.: Queries and concept learning, *Machine Learning*, Vol. 2, 1988, pp.319-342.

[3] Angluin, D.: Negative results for equivalence queries, *Machine Learning*, Vol. 5, 1990, pp.121-150.

[4] Balcazar, J. L., J. Diaz, R. Gavalda and O. Watanabe: A Note on the Query Complexity of Learning DFA, *Proc. of the 3rd workshop on Algorithmic Learning Theory*, Japanese Society for AI, 1992, pp.53-62.

[5] Benedek, G. M. and A. Itai: Nonuniform learnability, *International Colloquium on Automata, Languages and Programming 1988*, Lecture Notes in Computer Science 317, Springer-Varlag, pp.82-92.

[6] Blumer, A., A. Ehrenfeucht, D. Haussler, and M. K. Warmuth: Classifying learnable geometric concepts with Vapnik-Chervonenkis dimension, in *Proc. 18th ACM Symp. on Theory of Computing*, 1986, pp.273-282.

[7] Champarnaud, J.-M. and J.-E. Pin: A Maximin problem on finite automata, *Discrete Applied Mathematics 23*, 1989, pp.91-96.

[8] Hopcroft, J. E., and Ullman, J. D., *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley, Massachusetts (1979). (日本語訳: 野崎 昭弘他訳、オートマトン・言語理論・計算論 I, II (共立出版 1986))

[9] Ishigami, Y. and S. Tani: The VC-dimension of Finite Automata with n States, *Proc. of the 4th workshop on Algorithmic Learning Theory* (in Lecture Notes of Artificial Inteligence, Springer-Varlag) to appear.

[10] Gaizer, T.: The Vapnik-Chervonenkis dimension of finite automata. Unpublished manuscript (1990).

[11] Maass, T., and G. Turán : On the complexity of learning from counterexamples, in *Proc. 30th IEEE Symp. on Foundations of Computer Science* 1989, pp.262-167.

[12] Maass, T., and G. Turán: On the Complexity of Learning from Counterexamples and Membership Queries, in *Proc. 31st IEEE Symp. on Foundations of Computer Science*, 1990, pp.203-210.

[13] Pitt, L. and L. G. Valiant: Computational limitations on learning from examples, *Journal of the ACM*, 35 pp.965-984, 1988.

[14] Valiant, L.: Theory of the learnable. *Communications of the ACM*, 27(11) pp.1134-1142, 1984.

[15] Vapnik, V. N. and A. Chervonenkis: On the uniform convergence of relative frequencies, *Theory of Probability and its Applications*, 16 pp.264-280, 1971

[16] Watanabe, O.: A formal study of learnability via queries, *International Colloquium on Automata, Languages and Programming, 1990*, Lecture Notes in Computer Science 443, Springer-Varlag.

石上　嘉康 (Yoshiyasu Ishigami)
早稲田大学 理工学部数学科
169 東京都新宿区大久保 3-4-1
62m502@cfi.waseda.ac.jp

谷　聖一 (Sei'ichi Tani)
東京女子大学　情報処理教室
167 東京都杉並区善福寺 2-6-1
tani@twcu.ac.jp