

ストカステック言語

について

東北大・通研 那須正和 本多波雄

§ 1 序

有限オートマトンにおいて、初期状態の分布及び状態の遷移が確率的であるようなシステムを有限確率オートマトンという。すなわち、 λ 個のベクトル Σ 上の有限確率オートマトンとは、システム $\mathcal{A} = \langle S, \pi, \{A(i) | i \in \Sigma\}, \eta^F \rangle$ で与えられる。ここで S は有限個 (n 個) の状態の集合、 π は各要素が非負実数で n 個の要素の和が 1 である n 次元確率行ベクトルであり初期状態の分布を示す。 $A(i)$ ($i \in \Sigma$) は $n \times n$ 確率遷移行列でその (i, j) 要素は状態 s_i から状態 s_j へ遷移する確率を示す。 F は S の部分集合で最終状態の集合である。 η^F は n 次元列ベクトルで、その i 番目の要素 η_i は $s_i \in F$ ならば $\eta_i = 1$ 、 $s_i \notin F$ ならば $\eta_i = 0$ と定められる。

今 $x = \sigma_{i_1} \sigma_{i_2} \dots \sigma_{i_k}$ $\sigma_{i_j} \in \{E\} \cup \Sigma$ (E は空系列を示す) に対し、 $A(x) = A(\sigma_{i_1}) \dots A(\sigma_{i_k})$ (但し $A(E)$ は

マト Σ_0 が研究された。 $\Sigma_0 = \langle S, \pi, \{A(0), A(1)\}, \eta^F \rangle$,

$$S = \{s_1, s_2\}, \quad \pi = (1, 0) \quad A(0) = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$A(1) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix} \quad \eta^F = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{とされる。これによって受理}$$

される確率的事象を P とすると, $x = \sigma_1 \sigma_2 \dots \sigma_k$ $\sigma_i \in \{0, 1\}$

に対して, $P(x) = 0, \sigma_k \sigma_{k-1} \dots \sigma_1$ (2進展開) となることは容易にたしかめられる。

$0 \leq \lambda < \lambda_1 \leq 1$ に対して, $T(\Sigma_0, \lambda)$ は $T(\Sigma_0, \lambda_1)$ を真に含むから, この有限確率オートマトン

だけでなく, ストカスティック言語は実数の濃度だけ異

なるものが存在し, したがってストカスティック言語の内にはケ

ーリンツ"マツ"によって規定される言語の族に入らな

ものが存在することがわかる。そこで, ストカスティック言語でな

い言語はどのようなものかということが興味の対象となるが

しばらくの間 Bukharayev (1965) が recursive set の範囲内に

1-ツ"ボル"からなるアルファベットとでその例を見出したと伝

えられるにとどまっていた。しかし, コンテキスト・フリー

言語の内にもストカスティックでない言語は存在するであろうと予想

されていた。(例えば Salomaa (1969), 最近 Paz (1969)

が非常に簡明な方法によって, 1-ツ"ボル"アルファベットの

コンテキスト・センシティブ言語の内にはストカスティック言語でな

い言語の例を見出した。すなわち, $\Sigma = \{a\}$ とし, a, b 2ツ

シボルからなる全ての系列を辞書式にならべて、さらにそれらに順序をつないでできる無限系列 X において、 k 番目のシボルが σ であるとき、かつそのときに限り、 $\sigma^k \in L_S$ とすることによって L_S を定めると、 L_S は stochastic 言語ではない。このことを証明するためには、確率的事象の族 $\mathcal{P}$ が fuzzy event の族 $\mathcal{A}$ の真の部分集合であることを示す簡単な性質について述べておく。 P が確率的事象であると、任意の $x \in \Sigma^*$ に対して、 $C_0 + C_1 + \dots + C_n = 1$ ($C_n = 1$) なる有限個の実数の集合 $\{C_0, C_1, \dots, C_n\}$ が存在して、任意の $y, z \in \Sigma^*$ に対して、 $C_n P(yx^nz) + C_{n-1} P(yx^{n-1}z) + \dots + C_0 P(yz) = 0$ とすることがでる。このことは、 P がある有限確率オートマトン $\mathcal{A} = \langle S, \pi, \{A(\sigma) \mid \sigma \in \Sigma\}, \eta^F \rangle$ によって $P(yx^iz) = \pi A(y) A(x)^i A(z) \eta^F$ と書けることと、 $A(x)$ の characteristic polynomial を考えてみると容易にわかる。上に述べた性質を $P_{\mathcal{A}}$ に従って、finiteness property と呼ぶことにする。

L_S が stochastic 言語であるとすると、適当な確率的事象 P とカットポイント λ によって、 $L_S = \{x \in \Sigma^* \mid P(x) > \lambda\}$ とかける。finiteness property によって、任意の i に対して

$C_n + \dots + C_0 = 0$, $C_n P(\sigma^{n+i}) + C_{n-1} P(\sigma^{n+i-1}) + \dots + C_0 \cdot P(\sigma^i) = 0$ — (1) なる実数の組 $C = \{C_0, \dots, C_{n-1}\}$ が存在する。 C の内の正数を $C_{i_1}, C_{i_2}, \dots, C_{i_p}$ とする。 L_S の構成

法から明らかになるように、適当な λ をとれば、 $\sigma^L, \sigma^{L+1}, \dots, \sigma^{L+n}$ の内 $\sigma^{j+L}, \sigma^{j+L+1}, \dots, \sigma^{j+L+n}$ は L_S の元で、その他は L_S に含まれないようにすることができ、 $P(\sigma^{j+L}) > \lambda \Leftrightarrow \sigma^{j+L} \in L_S$ ($j=0, 1, \dots, n$) であるから、 $C_n P(\sigma^{n+L}) + C_{n-1} P(\sigma^{n+L-1}) + \dots + C_0 P(\sigma^L) = C_n (P(\sigma^{n+L}) - \lambda) + \dots + C_0 (P(\sigma^L) - \lambda) > 0$ となるので (i) と矛盾が生じる。したがって L_S はスタカスチック言語ではない。

同じ手法によって、スタカスチック言語ではないコネテフストフリー言語の存在を証明することが出来る。2-リミット・ボリ・リット $\{a, b\}$ 上で、 $a^i b a^{j_1} b \dots a^{j_r} b$ ($r=1, 2, \dots$) の形を持ち、適当な ℓ ($1 \leq \ell \leq r$) に対して、 $i = j_1 + j_2 + \dots + j_\ell$ が成立するような ℓ -ツプの集合を L_f とする。なお $i, j_1, \dots, j_r$ は非負整数とする。そうすると L_f は線形コネテフストフリー文法で生成されることを示される。このことにより、コネテフストフリー言語の内でも、正規集合の直接的拡張とみられる簡単な言語の範囲内にスタカスチック言語でないものが存在したことがわかった。後に述べるように、 L_f はスタカスチック言語の closure property に満たす事実を証明するのに有効である。

§ 3, closure property

有限確率オートマトンに関連して、2つのタイプの closure

Property の問題がある。1) は確率的事象の族 $\mathcal{F}$ の fuzzy set の意味での演算に関する問題であり、もう一つは、もっと広く、スタカスティック言語の族の演算に関して与える問題である。

$\mathcal{F}$ の closure property に関して次の結果が得られる。

- (1) $P, Q, r \in \mathcal{F}$ に対して, $(P, Q; r)(x) = P(x)r(x) + Q(x) \cdot (1-r(x))$ ($x \in \Sigma^*$) で与えられる fuzzy event $(P, Q; r)$ (コンビネース・コンビネーション) は確率的事象である。

(1) より (2) が導かれる。

- (2) (和集合, 共通集合), $P, Q \in \mathcal{F}$ に対して, fuzzy event $P \vee Q, P \wedge Q$ は $P \vee Q(x) = \max(P(x), Q(x)), P \wedge Q(x) = \min(P(x), Q(x))$ ($x \in \Sigma^*$) で定義すると, $P, Q \in \mathcal{F}$ であつて $\{x \in \Sigma^* \mid P(x) > Q(x)\}$ が正規集合ならば, $P \vee Q, P \wedge Q$ は共に確率的事象である。

この性質は次のように表現を変えれば, fuzzy set の概念を導入する意義がわかる。

- (2') 有限確率オートマトン $\mathcal{A}_1, \mathcal{A}_2$ が与えられ, $P, Q \in \mathcal{F}$ を受理するとし, $\{x \in \Sigma^* \mid P(x) > Q(x)\}$ が正規集合ならば, ある有限確率オートマトン $\mathcal{A}_3, \mathcal{A}_4$ が存在して, 任意のカットポイント $0 \leq \lambda \leq 1$ に対して, $T(\mathcal{A}_3, \lambda) = T(\mathcal{A}_1, \lambda) \cup T(\mathcal{A}_2, \lambda)$ $T(\mathcal{A}_4, \lambda) = T(\mathcal{A}_1, \lambda) \cap T(\mathcal{A}_2, \lambda)$ が成立する。

$P, Q \in \mathcal{F}$ に対して, $P \vee Q, P \wedge Q$ は一般的には $\mathcal{F}$ の元でな

これは, finiteness property を用いて, 松浦が最初に証明した。

(3) (逆事象) $P \in \mathcal{P}$, $P^T(x) \equiv P(x^T)$ ($x \in \Sigma^*$), 但し x^T は T - σ の逆にならべたものを示す。このとき $P^T \in \mathcal{P}$ 。

(3') 任意の有限確率オートマトン Σ に対して, ある有限確率オートマトン Σ_1 が存在して, 任意のカットポイント $0 \leq \lambda \leq 1$ に対して, $T(\Sigma, \lambda)^T = T(\Sigma_1, \lambda)$ が成立する。

(4) Σ , Δ をそれぞれアルファベットとする。 Σ^* から Δ^* への gsm 写像 Γ と Δ^* 上の確率的事象 P に対して, $\hat{\Gamma}(P)(x) = P(\Gamma(x))$ ($x \in \Sigma^*$) で与えられる Σ^* 上の fuzzy event $\hat{\Gamma}(P)$ は確率的事象である。

(4') Γ を Σ^* から Δ^* への gsm 写像とする, Δ 上の任意の有限確率オートマトン Σ_Δ に対して, ある Σ 上の有限確率オートマトン Σ_Σ が存在して, 任意のカットポイント $0 \leq \lambda \leq 1$ に対して $\Gamma^{-1}(T(\Sigma_\Delta, \lambda)) = T(\Sigma_\Sigma, \lambda)$ が成立する。

有限確率オートマトン $\Sigma = \langle S, \pi, \{A(i) \mid i \in \Sigma\}, \eta^F \rangle$ において, $A(i), i \in \Sigma$, π の各要素が全て有理数とすると, Σ を有理数有限確率オートマトン (RFPA) という。RFPA によって受理される確率的事象を有理数確率的事象 (RPE) ということにし, その族を $\mathcal{P}_R$ とする。そうすると $P, Q \in \mathcal{P}_R$ に対して, $\{x \in \Sigma^* \mid P(x) = Q(x)\}$ 及び $\{x \in \Sigma^* \mid P(x) \neq Q(x)\}$

はストカス $\bar{T}_1$ ツク言語となる。(松浦, 等, Turakainen 1968)
 $\{x \in \Sigma^* \mid p(x) = q(x) \mid p, q \in \mathcal{P}_R\}$ なる形の言語を E-言語,
 $\{x \in \Sigma^* \mid p(x) \neq q(x) \mid p, q \in \mathcal{P}_R\}$ なる形の言語を D-言語,
 $\{x \in \Sigma^* \mid p(x) > q(x) \mid p, q \in \mathcal{P}_R\}$ なる形の言語を P-言語とそ
 れぞれいうことにすれば, P 言語はコ- $\bar{T}_1$ + スト-セ- $\bar{T}_1$ フ
 言語に真に含まれる(郡倉, 等 1968)。E 言語でも D 言語
 でもなる P 言語は存在する(例えは $\{a^m b^n \mid m \geq 0, n \geq 0, m > n\}$)
) けれども, E 言語と D 言語はもとより具体的に作
 り出せるストカス $\bar{T}_1$ ツク言語である。前の性質(4)を用いて, 任意
 の Σ^* から Δ^* への gsm 写像 Γ_1, Γ_2 に対して, $\{x \in \Sigma^* \mid \Gamma_1(x) = \Gamma_2(x)\}$
 $\{x \in \Sigma^* \mid \Gamma_1(x) \neq \Gamma_2(x)\}$ なる形の言語はストカス $\bar{T}_1$ ツク言
 語であることがわかり, 又同じ考え方に依り, 次のような

規則 P を持つ文法 (V, Σ, P, σ) (d.l 文法という) で生
 成される言語 (d.l 言語という) はストカス $\bar{T}_1$ ツク言語である
 ことがいえる。(i) $v \rightarrow a \xi u$ あるいは $v \rightarrow b$, $v, \xi \in V - \Sigma$
 $a, b \in \Sigma$, $u \in \Sigma^*$ の形を持ち (ii) 任意の 2 つの規則 $v_1 \rightarrow a_1 x$
 $v_2 \rightarrow a_2 y$, $a_1, a_2 \in \Sigma$, $x, y \in \{\varepsilon\} \cup (V - \Sigma) \Sigma^*$, $v_1, v_2 \in V - \Sigma$
 に対して, $v_1 = v_2$ かつ $a_1 = a_2$ ならば $x = y$ がなりたつ。

$L_1 = \{a^i b a^{j_1} \dots b a^{j_r} \mid i, j_1, \dots, j_r \geq 0, i = j_1 + \dots + j_r\}$ は
 E 言語したがってストカス $\bar{T}_1$ ツク言語である (Turakainen)。正
 規集合 $b(\varepsilon \cup \{a, b\}^* b) \in L_2$ とすると, $L_1 L_2 = L_f$ となる

から、ストカステック言語の族は、 Γ -カテゴリー- $\mathcal{L}_3$ の演算のもとで閉じていない。又 $L_3 = L_1 C(\varepsilon \cup \{a, b\}^* b)$ はストカステック言語であることが証明できる。そこで $h(a) = a, h(b) = b, h(c) = b$ なるホモモルフィズムを考えると、 $h(L_3) = L_f$ となるから、ストカステック言語の族はホモモルフィズムの演算のもとで閉じていない。Tsurukainen は L_f を与えられて、我々と同様に上の結果を得、さらに $L' = \{a^k b (a^* b)^* a^k b \mid k \geq 0\}$ がストカステック言語であることを示し、 L'^* がストカステック言語でないことを同様の手法で証明した。

6.3. いくつかの決定不可能問題

最後に RFPA に属する、3 の決定問題について述べる。

以下の事実は、E-言語によって、ストカステック言語を構成し、Post の Correspondence problem に帰着させることによって証明できる。

(1) 任意の RFPA Σ と有理数 λ に対して、 $T(\Sigma, \lambda)$ は空か否か、 $\emptyset$, $T(\Sigma, \lambda)$ は Σ^* か否か、 $T(\Sigma, \lambda)$ は正規集合であるかどうか、 $T(\Sigma, \lambda)$ はコンテキスト・フリーであるかどうかの4つの問題は全て帰納的に決定不可能である。

(2) 任意の RFPA Σ_1 と Σ_2 に対して、それが受理する確率的事象をそれぞれ P_1, P_2 とすると、 $1, P_1 \vee P_2$ は確率的事象で

あるか 口, $P_1 \wedge P_2$ は確率的事象であるか の 2 つの問題は
帰納的に決定不可能である。

参考文献

- Bar-Helll, Y., Perles, M. and Shamir, E.: On formal properties of
simple phrase structure grammars, Z. Phonetik, Sprach, Kommun-
-icationstorsch., vol. 14 (1961) 143-
- Bukharuev, R.: Criteria for the representation of events in finite
probabilistic automata. Dokl. Akad. Nauk SSSR 164 (1965)
- Gensburg, S.: Mathematical Theory of Context Free Languages. McGraw
-Hill, New York (1966)
- Harrison, M. A.: Introduction to Switching and Automata Theory
McGraw-Hill, New York (1965)
- 松浦, 穂垣, 福村.: オートマトンの一般化とその解析 オートマトン
研究会資料 (1968-1)
- 松浦, 穂垣, 福村.: 線形空間オートマトンと確率オートマトン.
電子通信学会全国大会講演論文集 S8-4 昭和43年10月
- Nasu, M. and Honda, N. Fuzzy events realized by finite probabilistic
automaton Information and Control 12 284-303 (1968)
- Nasu, M and Honda, N. Mappings induced by PGSM-mappings
and some recursively unsolvable problems of finite probabilistic

- automata. Information and Control 15 250-272 (1969)
- Nasu, M. and Honda, N. A context free language which is not acceptable by a probabilistic automaton. Information and Control 掲載決定 (1971)
- Paz, A. Some aspects of probabilistic automata. Information and Control 9 26-60 (1966)
- Paz, A. Events which are not representable by a probabilistic automaton (preliminary draft) (1969)
- Paz, A. Formal Series, Finiteness properties and Decision Problems Technical Report NO.4 ISRAEL INSTITUTE OF TECHNOLOGY. (1970)
- Rabin, M.O. Probabilistic Automata, Information and Control 6 230-245 (1963)
- Salomaa, A. Theory of Automata Pergamon Press Oxford (1969)
- Salomaa, A., Probabilistic and weighted grammars. Information and Control 15 529-544 (1969)
- Schützenberger, M.P. Certain elementary families of automata, Symposium on Mathematical theory of Automata, Polytechnic Institute of Brooklyn (1962)
- 都倉, 藤井, 著: 線形オートマトンに用いる3, 3の考察
電子通信学会全国大会講演論文集 58-2 昭和43年10月
- Turakainen, P. On probabilistic automata and their genera

-izations. ANNALS ACADEMIE SCIENTIARUM FENICAE

Serie A I MATHEMATICA 429 (1968)

Turakainen, P., The family of stochastic languages is closed
neither under catenation nor under homomorphism.

Ann. Univ. Turku, A I 133 (1970)

Turakainen, P., Some closure properties of the family of
stochastic languages, manuscript, (1970)

Zadeh, A.L., Fuzzy sets. Information and Control 8
338-353 (1965)