

メモリ量を低減した近似ベイズ符号化アルゴリズム

野村 亮[†] 松嶋 敏泰[†] 平澤 茂一[†]

An Approximation Algorithm of Bayes Coding to Reduce Memory Capacity

Ryo NOMURA[†], Toshiyasu MATSUSHIMA[†], and Shigeichi HIRASAWA[†]

あらまし 情報源の確率モデルは既知であるが、そのパラメータは未知である場合の符号化法において、ベイズ符号はベイズ基準のもとで冗長度を最小にする符号である。また、ベイズ符号を構成するアルゴリズムとして FSMX 情報源に対する文脈木を用いたベイズ符号化法が提案されている。このアルゴリズムは文脈木を逐次的に生成することにより、最大深さが任意の FSMX 情報源に対してベイズ符号を構成している。しかし、実用化を考えた場合、系列長とともに文脈木を生成することはメモリの点から困難である。本論文ではメモリ容量を低減した近似ベイズ符号化アルゴリズムを提案しその性能を評価する。

キーワード ベイズ符号, ユニバーサル符号, 文脈木, 事後分布

1. ま え が き

情報源の確率構造について完全な情報が得られていない場合の符号化法、すなわちユニバーサル情報源符号化法に関しては従来より多くの研究がなされている [1], [2]。情報源の分布のクラスのみを仮定し、そのパラメータに関しては未知の場合を扱うユニバーサル符号の中で、ベイズ符号は冗長度をベイズ基準のもとで最小にする符号である [3]。そして、FSMX 情報源に対して文脈木を用いたベイズ符号の効率的なアルゴリズムが提案されている [4], [5]。これらのアルゴリズムはあらかじめ文脈木の深さを設定しており、FSMX 情報源の一部のクラスに対するベイズ符号を構成するものであった。一方近年、文脈木を逐次的に生成することによりすべての FSMX 情報源のクラスを対象としたベイズ符号化アルゴリズムが提案されている [6], [7]。しかし、これらのアルゴリズムにおいては文脈木のノード数が系列長とともに増大するため、実用化の際には莫大なメモリ量を必要としてしまう。情報源の確率構造を全く仮定しないユニバーサル符号である Ziv-Lempel (LZ) 符号 [8] は漸的に最適な符号であるが、実は同様の問題を抱えており、実用化の際にはある深さまでしか文脈木を成長させない、な

どのアルゴリズムでこの問題を回避している。

本論文では、文脈木のノード数をメモリ量と考え、まずノード数を低減したアルゴリズムを提案する。次に、提案アルゴリズムを用いたときの 1 シンボル当りの符号長がベイズ符号化法の符号長と漸的に一致することを示す。更に、提案アルゴリズムの有限時点での性能をいくつかの数値実験により評価する。

2. ベイズ符号化法

2.1 FSMX 情報源

ベイズ符号は情報源の分布のクラスのみが既知であり、そのパラメータが未知の場合を対象とした符号である。松嶋らにより提案されたベイズ符号を構成する効率的なアルゴリズム(以下、ベイズ符号化法と呼ぶ)は有限アルファベット上の FSMX 情報源を対象としている。次に FSMX 情報源について定式化を行う。

FSMX 情報源とは過去の有限系列から現在のシンボルの発生確率の決まる情報源でマルコフ過程の一種である。FSMX 情報源は階層型モデルであるため、モデルとそのモデルのもとでのパラメータの二つにより定まる。 $x^n : x_1 x_2 \cdots x_n$ を長さ n の情報源系列とすると、FSMX 情報源における t 時点の状態は情報源系列 x^{t-1} により決まる。FSMX 情報源モデル m における状態の集合を $S(m)$ であるとし、この情報源系列 x^{t-1} から状態 $s \in S(m)$ への写像を $s(x^{t-1})$ とする。ここで、情報源アルファベットを

[†] 早稲田大学理工学部経営システム工学科, 東京都
School of Science and Engineering, Waseda University, 3-4-
1 Okubo, Shinjuku-ku, Tokyo, 169-8555 Japan

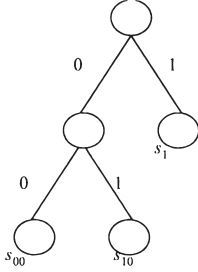


図 1 binary-FSMX 情報源モデル (木表現)
Fig.1 An example of binary-FSMX source model.

$a \in A = \{a | 0 \leq a \leq l-1\}$ とすると各状態 s でのシンボルの出現確率は $(l-1)$ 次元パラメータベクトル $\theta^s = \{\theta_1^s, \theta_2^s, \dots, \theta_{l-1}^s\}$ によって決まる。したがって、シンボル x_t の x^{t-1} のもとでの条件付き確率は $P(x_t | \theta^s(x^{t-1}), s(x^{t-1}))$ となる。

結局、情報源系列 x^t の発生確率は以下で表される。

$$P(x^t) = P(x_1 | \theta^{s(\lambda)}, s(\lambda)) P(x_2 | \theta^{s(x^1)}, s(x^1)) \dots P(x_t | \theta^{s(x^{t-1})}, s(x^{t-1})), \quad (1)$$

ここで、 $s(\lambda)$ は初期状態を表す。

$l-1$ 次元パラメータベクトル θ^s は各状態 s に対応しているので一つの FSMX 情報源のパラメータを $|S(m)|(l-1)$ 次元パラメータベクトル θ^m で表すことにする。すると、FSMX 情報源はモデル m とそのもとでのパラメータを表す θ^m により定義される。

FSMX 情報源モデル m はまた、完全木で表現することができる。木におけるそれぞれの枝はシンボル $a \in A$ に対応している。また、木における葉ノードから根ノードへの一つのパスをコンテキスト若しくはポストフィクスと呼ぶ。木表現におけるそれぞれの葉ノードは状態 s と 1 対 1 対応しているので、葉ノードは s と書くことができ、FSMX 情報源モデル m における状態の集合 $S(m)$ は木表現における葉ノードの集合といえる。

図 1 に 2 元 FSMX 情報源モデル m_1 の木表現例を示す。文脈 $x^{t-1} = \dots 10$ で決定される状態を s_{10} と書くことにすると、系列 $x^5 = 10010$ において $t = 2$ 時点の状態は $s(1) = s_1$ 、 $t = 3$ 時点の状態は $s(10) = s_{10}$ 、 $t = 4$ 時点の状態は $s(100) = s_{00}$ となる。また、 $S(m_1) = \{s_1, s_{10}, s_{00}\}$ である。

なお、本論文では 2 元 FSMX 情報源 ($l = 2$) を考え、パラメータのとり得る範囲は 0 から 1 の開区間、すなわち $\theta^s \in (0, 1)$ 、とする。

2.2 FSMX 情報源に対するベイズ符号化法

情報源系列の確率を仮定すれば算術符号を用いることにより符号化が可能であるため、ユニバーサル情報源符号化の問題はシンボルの出現確率を決定する問題に帰着する。ベイズ符号はベイズ基準のもとで冗長度を最小にシンボルの出現確率 (以下、符号化確率と呼ぶ) を決定する。FSMX 情報源モデル m とそのもとでのパラメータ θ^m が共に未知でありかつ、FSMX 情報源モデル m の事前確率 $P(m)$ 、モデル m のもとでのパラメータ θ^m の事前確率 $P(\theta^m | m)$ が既知であるとき、ベイズ符号の符号化確率は以下の式で求められる。

[補題 2.1][3] FSMX 情報源 (m, θ^m) に対するベイズ符号の符号化確率は

$$AP(x_t | x^{t-1}) = \sum_{m \in M} \int_{\theta^m} P(x_t | x^{t-1}, \theta^m, m) \cdot P(\theta^m | m, x^{t-1}) P(m | x^{t-1}) d\theta^m, \quad (2)$$

である。ここで、 $P(\theta^m | m, x^{t-1})$ は FSMX 情報源 m と x^{t-1} のもとでの θ^m の事後確率、また $P(m | x^{t-1})$ は x^{t-1} のもとでの m の事後確率を表す。□

ベイズ符号化法は上記の符号化確率を効率的に求めるアルゴリズムである。以下、そのアルゴリズムについて説明する。ベイズ符号化法は各時点ごとに 1) 文脈木の生成を行い、その文脈木を用いて 2) 符号化確率の計算を行う。最初に、文脈木の生成法について説明する。

2.2.1 文脈木の生成

部分系列 x_i^j を $x_i^j = x_i x_{i+1} \dots x_j$ とする。 t 時点のシンボル x_t に対してそのポストフィクス^(注1)に対応するノード $s(\lambda), s(x_{t-1}^t), s(x_{t-2}^t), \dots, s(x_1^t)$ のうち、既に文脈木に含まれているノードを除いたノードを文脈木に加える。□

t 時点のポストフィクスに対応するノードの集合を S_t と書くことにする。つまり、 $S_t = \{s(\lambda), s(x_{t-1}^t), s(x_{t-2}^t), \dots, s(x_1^t)\}$ となる。

このようにして生成された文脈木は現在得られている系列から考えられるすべての FSMX 情報源の状態の集合となっている。

更に、松嶋らは以下の式を提案した [5]。

(注1): $x^5 = 01001$ のとき、 $x_5 (= 1)$ のポストフィクスは、部分系列の集合 $\{\lambda, 0, 00, 100, 0100\}$ である。ここで λ は空系列を表す。

$$P(s) = \sum_{\{m|s \in S(m)\}} P(m). \quad (3)$$

[注意 2.1] ここで、上記の確率は次の式を満たしている。

$$\sum_{s \in S_t} P(s) = 1. \quad (4)$$

この式は文脈木における一つのノードの事前確率にそのノードを状態として含む FSMX 情報源モデルの事前確率を対応させたものである。この式を用いることにより、一つのポストフィクスに属するノードを考えるのみでベイズ符号の符号化確率を計算することができる。式 (3) を用い、更にパラメータの事前分布にディレクレ分布を仮定すると、符号化確率の計算式は次の補題で与えられる。

2.2.2 符号化確率の計算

[補題 2.2] [7]

$$\begin{aligned} AP(x_t|x^{t-1}) \\ = \sum_{s \in S_t} P^s(x_t|x^{t-1}, s)P(s|x^{t-1}), \end{aligned} \quad (5)$$

ここで、

$$P^s(x_t|x^{t-1}, s) = \frac{n(x_t|x^{t-1}, s) + \beta(x_t|s)}{\sum_{i=0}^{t-1} n(i|x^{t-1}, s) + \beta(s)}, \quad (6)$$

である。また、 $\beta(x_t|s)$ 、 $\beta(s)$ はそれぞれ既知のディレクレ分布のパラメータである。また $n(i|x^{t-1}, s)$ は x^{t-1} における状態 s のもとでのシンボル i の発生回数で x^{t-1} により計算される。□

式 (5) における和は t 時点のポストフィクスに対してとられる。ゆえに、文脈木と式 (5) を用いれば、符号化確率を計算する際に、 t 時点のポストフィクスのみを考えればよいことになる。

ここで、ノードの事後確率は以下のように更新される。

$$P(s|x^t) = \frac{q(s|x^t)}{\sum_{s \in S_t} q(s|x^t)} \quad (7)$$

ただし、

$$\begin{aligned} q(s|x^t) \\ = \begin{cases} \frac{P^s(x_t|x^{t-1}, s)P(s|x^{t-1})}{AP(x_t|x^{t-1})}, & \text{if } s \in S_t \\ P(s|x^{t-1}), & \text{if } s \notin S_t \end{cases} \end{aligned} \quad (8)$$

である。また、 $P(s|x^0) = P(s)$ は事前分布であり、既

知の $P(m)$ より計算される。

上記のもとで、次の式が成り立つことに注意されたい。

$$P(s|x^t) = \sum_{\{m|s \in S(m)\}} P(m|x^t). \quad (9)$$

松嶋らにより上記の更新過程の簡略法が提案されている [5]。

2.3 ベイズ符号化法の問題点

前節で述べたとおり、松嶋らにより提案されたベイズ符号化法を用いることにより、符号化確率の計算は非常に効率的に行うことができる。しかし、文脈木の生成において系列長が増加すればするほど、文脈木の深さも増加してしまう。

メモリ容量の面から増加するすべてのノードを保持しておくことは非常に困難である。そのため、ベイズ符号はベイズ基準のもとで最適な符号であり、その構成アルゴリズムも提案されているが、このままでは実用化は困難と考えられる。

3. 提案アルゴリズム

3.1 FSMX 情報源に対するベイズ符号の性質

提案アルゴリズムの前に、本節で FSMX 情報源の性質と FSMX 情報源に対するベイズ符号化法の性質を調査する。

FSMX 情報源 (m^* 、 θ^{*m^*}) によりデータが発生しているとする。本研究においてこれらをそれぞれ真のモデル m^* 、真のモデルにおける真のパラメータ θ^{*m^*} と呼ぶことにする。更に、 m^* の要素を s^* とし、真のノードと呼ぶことにする。真のモデルにおける真のパラメータ θ^{*m^*} を真のノードごとに対応させた θ^{*s^*} を真のノードにおける真のパラメータと呼ぶことにする。更に、ノード s' が文脈木においてノード s の子孫にあたる場合、その関係を $s' \gg s$ と書くことにする。

ここで、真のノード s^* の定常分布を $q(s^*)$ と書くことにし、文脈木における各ノード s に対して次の値を定義する。

$$\theta^{*s} = \begin{cases} \theta^{*s^*} & \text{if } s \gg s^* \\ \frac{\sum_{s^* \in \bar{S}^*} q(s^*)\theta^{*s^*}}{\sum_{s^* \in \bar{S}^*} q(s^*)} & \text{if } s \ll s^*. \end{cases}$$

ここで、 \bar{S}^* は真のノードの集合の中でノード s の子孫にあたるノードの集合を指す。すなわち、 $\bar{S}^* = \{s^* : s^* \gg s\}$ である。上記により定義されるパラメータを各ノードにおける真のパラメータと呼

ぶことにする．つまり，真のノードより子孫側にあればそのノードの真のパラメータは真のノードのそれと同じである．また，真のノードより親側にあるノードに対しては，その子孫側に真のノードが複数存在するので，それらを定常分布で荷重平均してやったものを真のパラメータと呼ぶ．これらを各ノードの真のパラメータと呼ぶのは以下が成り立つことによる．

$\forall i \in A$ に対して

$$\lim_{n \rightarrow \infty} \frac{n(i|x^n, s)}{\sum_{i=0}^{l-1} n(i|x^n, s)} = \theta_i^{*s} \quad a.s., \quad (10)$$

となる．上式は各ノードにおいてシンボル i の出現頻度が θ_i^{*s} に概収束することを意味する．これより文脈木における任意の FSMX 情報源モデル m に対して，真のパラメータ， $\theta^{*m} = \{\theta^{*s_{i1}}, \dots, \theta^{*s_{i|S(m)|}}\}$ ，ここで $s_{i1}, \dots, s_{i|S(m)|} \in S(m)$ ，が定義できる．

更に仮定の前に次を準備する．深さ J の葉ノード s^J をもつ FSMX 情報源モデル m_{s^J} に対して， $S(m_{s^J}) = \{s_{i1}, \dots, s_{i(|S(m)|-1)}, s^J\}$ であるとし， $J \neq 0$ である場合を考える．ノード s^J に対して，木表現におけるその親ノードを s^{J-1} とすると，2元 FSMX 情報源モデルを考えているため s^{J-1} を祖先としてもつ葉ノードが $S(m_{s^J})$ 内に s^J 以外に一つ以上存在する．このノードの集合を S^J とすると $S(m_{s^{J-1}}) = (S(m_{s^J}) \cup \{s^{J-1}\}) \setminus (\{s^J\} \cup S^J)$ なる FSMX 情報源モデル $m_{s^{J-1}}$ が定義できる．また同様にノード s^{J-1} に対しても $s^{J-1} \neq s_\lambda$ であれば，FSMX 情報源モデル $m_{s^{J-2}}$ が定義できる．このように一つの FSMX 情報源モデルに対して，ある深さ J の葉ノード s^J に注目し上記の操作を繰り返すと $J+1$ 個の FSMX 情報源モデルが定義できる．すると s^J, \dots, s^0 はポストフィクスをなす．すなわち， $s^0 = s_\lambda$ であり， $s^0 \ll s^2 \ll \dots \ll s^J$ が成立している．次に仮定を述べる．

[仮定 3.1] 葉ノード s^J をもつある FSMX 情報源に対して， $0 \leq j < i \leq J$ とすると

$$D(P_{m^*}^*; P_{m_{s^j}^*}^*) - D(P_{m^*}^*; P_{m_{s^i}^*}^*) > 0, \quad (11)$$

が成立する．ここで， P_m^* は確率分布 $P(X|\theta^{*m}, m)$ を表す．また， $D(P(X); Q(X))$ は確率分布 $P(X)$ と $Q(X)$ の間の KL 情報量である． □

上記の仮定は，ある FSMX 情報源と KL 情報量的に最も近い任意の低次の FSMX 情報源に対して，同様の確率分布を表現可能なより低次の FSMX 情報源

が存在しないことを意味している．

[例 3.1] 深さ 2 の 2 元マルコフ情報源を考える．各状態でのシンボルの出現確率が次のような場合，この情報源は仮定 3.1 を満たす． $P(0|00) = 0.2, P(0|10) = 0.3, P(0|01) = 0.4, P(0|11) = 0.5$.

また，最大深さ 2 の FSMX 情報源の各状態でのシンボルの出現確率が次のような場合を考える． $P(0|0) = 0.5, P(0|01) = 0.4, P(0|11) = 0.6$. ここで， $s^2 = s_{11}$ であるとする， $s^1 = s_1, s^0 = s_\lambda$ である．ここで， $S(m_{s^1}) = \{s_0, s_1\}$ であり， $S(m_{s^0}) = \{s_\lambda\}$ である．また， $P(0|\theta^{*s_0}, s_0) = P(0|\theta^{*s_1}, s_1) = 0.5$ ， $P(0|\theta^{*s_\lambda}, s_\lambda) = 0.5$ である．ゆえに，

$$D(P_{m^*}^*; P_{m_{s^1}^*}^*) - D(P_{m^*}^*; P_{m_{s^0}^*}^*) = 0, \quad (12)$$

となるのでこの情報源は仮定 3.1 を満たさない． □

FSMX 情報源に対しては以下の式が成立することが示されている [13] .

[性質 3.1] 重複対数の法則が成り立つ．任意の i に対して，

$$\limsup_{n \rightarrow \infty} \frac{n(i|x^n, s) - n\theta_i^{*s}}{(2n\sigma \log \log n)^{\frac{1}{2}}} = 1 \quad a.s. \quad (13)$$

σ は FSMX 情報源におけるシンボルの分散とする．上記の性質より次が成り立つ．

[系 3.1] $\hat{\theta}^m(x^n)$ を θ^m の最ゆう推定量とすると，

$$\|\hat{\theta}^m(x^n) - \theta^{*m}\| \leq O\left(\left(\frac{\log \log n}{n}\right)^{\frac{1}{2}}\right) \quad a.s., \quad (14)$$

が成立する． □

[性質 3.2] パラメータの最ゆう推定量が以下を満たす．

$$P(\hat{\theta}^m(x^n)|m, x^n) = \left(\frac{n}{2\pi}\right)^{\frac{km}{2}} \sqrt{\det I(\hat{\theta}^m(x^n)|m)} + o(n^{-\frac{km}{2}}) \quad a.s., \quad (15)$$

ここで， $I(\theta^m|m)$ は，パラメータ θ^m の Fisher 情報量行列を表す．すなわち，

$$I(\theta^m|m) = - \lim_{n \rightarrow \infty} \frac{1}{n} E \frac{\partial^2 \log P(x^n|m, \theta^m)}{\partial \theta^m \partial \theta^{mT}}, \quad (16)$$

である． □

また，大数の強法則より以下も成り立つ．

[性質 3.3] $\forall \theta^m, m$ に対して

$$\begin{aligned} & \frac{1}{n} \log \frac{P(x^n | m^*, \theta^{m^*})}{P(x^n | m, \theta^m)} - D(p_{m^*}^{\theta^*} \| p_m^{\theta^m}) \\ & = o(1) \text{ a.s.}, \end{aligned} \quad (17)$$

となる。ここで、 $D(p_{m^*}^{\theta^*} \| p_m^{\theta^m})$ は分布 $p(X|\theta^{m^*}, m^*)$ と分布 $p(X|\theta^m, m)$ の間の KL 情報量を示す。□

上記のもとで後藤らは次の補題を示した [13]。

[補題 3.1] [13] 任意の FSMX 情報源 $m \neq m^*$ に対して、適当な事前分布のもとで、

$$\lim_{n \rightarrow \infty} \left| \frac{P(x^n | m)}{P(x^n | m^*)} \right| = 0 \text{ a.s.}, \quad (18)$$

が成立する。ただし、 m の集合 \mathcal{M} は有限集合。

ここで、上記補題の証明に集合 \mathcal{M} は有限である条件を用いていないことより、次の系が成立することに注意されたい。

[系 3.2] 任意の FSMX 情報源 $m \neq m^*$ に対して、適当な事前分布のもとで、

$$\lim_{n \rightarrow \infty} \left| \frac{P(x^n | m)}{P(x^n | m^*)} \right| = 0 \text{ a.s.}, \quad (19)$$

が成立する。ただし、 m の集合 \mathcal{M} は可算無限集合。□

また、次の補題も成立する。

[補題 3.2] $s' \ll s^*$ (若しくは、 $s^* \ll s'$) を満たす s' について、適当な事前分布のもとで、

$$\lim_{n \rightarrow \infty} \left| \frac{P(x^n | m_{s'})}{P(x^n | m_s)} \right| = 0 \text{ a.s.}, \quad (20)$$

を満たす $s : s' \ll s \ll s^* (s^* \ll s \ll s')$ が存在する。ただし $s \ll s'$ は、 $s \ll s'$ または $s = s'$ を意味することとする。

(証明) 付録参照。□

また、次を仮定する

[仮定 3.2] 任意の s に対してその子ノードを s^1 としたとき

$$P(s^1) = \frac{1}{r} P(s), \quad (21)$$

ここで $r > 1$ である。□

上記の仮定は Willems らによる CTW 法においても仮定されている [4]。

これらの系、補題より、FSMX 情報源におけるベイズ符号の漸近的性質に関する定理を得る。ただし、 S を n 時点の文脈木における任意のポストフィクスとする。

[定理 3.1] 任意の FSMX 情報源と任意の $s \neq s^*$ となる $s, s^* \in S$ に対して、

$$\lim_{n \rightarrow \infty} \left| \frac{P(s|x^n)}{P(s^*|x^n)} \right| = 0 \text{ a.s.} \quad (22)$$

が成り立つ。

(証明) 式 (9) より、

$$\begin{aligned} & \frac{P(s|x^n)}{P(s^*|x^n)} \\ & = \frac{\sum_{\{m|s \in S(m)\}} P(x^n|m)P(m)}{\sum_{\{m|s^* \in S(m)\}} P(x^n|m)P(m)} \end{aligned} \quad (23)$$

$$\leq \frac{\sum_{\{m|s \in S(m)\}} P(x^n|m)P(m)}{P(x^n|m^*)P(m^*)}, \quad (24)$$

である。上式分子の集合は可算無限個の要素をもつ。そこで、次のように分けて考える。

$$\bar{M}_s = \{m | D(P_{m^*}^*; P_m^*) > 0\}, \quad (25)$$

$$\bar{M}_s^C = \{m | D(P_{m^*}^*; P_m^*) = 0\}. \quad (26)$$

ここで、補題 3.2 の証明より、任意の $m \in \bar{M}_s^C$ に対して

$$\lim_{n \rightarrow \infty} \frac{p(x^n|m)}{p(x^n|m^*)} = O\left(\frac{1}{n^{\frac{k_m^* - k_m}{2}}}\right) \quad (27)$$

である。また、任意の $m \in \bar{M}_s^C$ において、 $k_m > k_m^*$ である。ゆえに c_i を、 \bar{M}_s^C において $k_m = k_m^* + i$ であるモデルの数とすると

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\sum_{m \in \bar{M}_s^C} P(x^n|m)P(m)}{P(x^n|m^*)P(m^*)} \\ & = \sum_{i=1}^{\infty} O\left(\frac{c_i}{n^{\frac{i}{2}}}\right) \text{ a.s.}, \end{aligned} \quad (28)$$

である。ここで、 $c_i \leq |S(m^*)|^i$ であることを考えると、結局

$$\sum_{i=1}^{\infty} O\left(\frac{c_i}{n^{\frac{i}{2}}}\right) \leq \sum_{i=1}^{\infty} O\left(\frac{|S(m^*)|^i}{n^{\frac{i}{2}}}\right), \quad (29)$$

が成立する。ここで、

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{|S(m^*)|^i}{n^{\frac{i}{2}}} & = \sum_{i=1}^{\infty} \left(\frac{|S(m^*)|}{n^{\frac{1}{2}}}\right)^i \\ & = \sum_{i=1}^{\infty} \left(\frac{|S(m^*)|}{n^{\frac{1}{2}}}\right)^{i-1} - 1, \end{aligned} \quad (30)$$

である．ここで $\sum_{i=1}^{\infty} \left(\frac{|S(m^*)|}{n^{\frac{1}{2}}}\right)^{i-1}$ は初項 1, 項比 $\frac{|S(m^*)|}{n^{\frac{1}{2}}}$ の等比級数で, その和は n が十分大きいとき, 1 に収束する．ゆえに

$$\sum_{i=1}^{\infty} \left(\frac{|S(m^*)|}{n^{\frac{1}{2}}}\right)^{i-1} - 1 = 0, \quad (31)$$

となる．上式は式 (28) が 0 に収束することを示している．

一方, \bar{M}_s は $D(P_{m^*}; P_m^*)$ の値により, 更にいくつかの集合に分割することができる．

例えば,

$$S(m^*) = \{s(000), s(100), s(10), s(01), s(011), s(111)\} \quad (32)$$

であるとし, $s^* = s(000), s = s(00)$, であるとする．この場合, $S(m) = \{s(00), s(10), s(1)\}$, というモデルと同様の確率分布を表現する集合と, $S(m) = \{s(00), s(10), s(01), s(11)\}$, というモデルと同様の確率分布を表現する集合, $\{s(00), s(10), s(01), s(011), s(111)\}$, と同様の確率分布を表現する集合の三つに分けることができる．

この集合を $\bar{M}_{s_1}, \bar{M}_{s_2}, \dots, \bar{M}_{s_j}$, とすると, j は m^* と s に依存するが有限である．ここで, 補題 3.2 の証明より, 任意の $m \in \bar{s}_i$ に対して

$$\lim_{n \rightarrow \infty} \frac{p(x^n|m)}{p(x^n|m^*)} = O(e^{-nD(P_{m^*}; P_m^*)}) \quad a.s., \quad (33)$$

であるので,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\sum_{m \in \bar{M}_{s_i}} P(x^n|m)P(m)}{P(x^n|m^*)P(m^*)} \\ &= \sum_{i=1}^{\infty} O\left(\frac{|S(m^*)|^i}{e^{nD(P_{m^*}; P_m^*)}}\right) \quad a.s., \quad (34) \end{aligned}$$

が成立する．上式は 0 に収束することがわかるので, 結局

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{\sum_{\{m|s \in S(m)\}} P(x^n|m)P(m)}{P(x^n|m^*)P(m^*)} \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^j \frac{\sum_{m \in \bar{M}_{s_i}} P(x^n|m)P(m)}{P(x^n|m^*)P(m^*)} \\ & \quad + \frac{\sum_{m \in \bar{M}_s^C} P(x^n|m)P(m)}{P(x^n|m^*)P(m^*)} \\ &= 0 \quad a.s., \quad (35) \end{aligned}$$

を得る．これは定理が成り立つことを示している． \square

以上の議論より以下の補題が成り立つ．

[補題 3.3] $\forall s^* \in S(m)$ に対して仮定 3.2 のもとで

$$\lim_{t \rightarrow \infty} |P^s(x_t|x^{t-1}, s^*(x^{t-1})) - AP(x_t|x^{t-1})| = 0 \quad a.s., \quad (36)$$

が成立する．ここで, $s^*(x^{t-1})$ は x^{t-1} のポストフィクスにおける s^* を表す． \square

(証明)

$$\begin{aligned} & \sum_{s: s \neq s^*, s \in S_n} \frac{P(s|x^n)}{P(s^*|x^n)} \\ &= \sum_{s: s \neq s^*, s \ll s^*, s \in S_n} \frac{P(s|x^n)}{P(s^*|x^n)} \\ & \quad + \sum_{s: s \neq s^*, s \gg s^*, s \in S_n} \frac{P(s|x^n)}{P(s^*|x^n)} \quad (37) \end{aligned}$$

が 0 に収束することを示す．ここで, 上式の右辺第 1 項は有限個の和であるので定理 3.1 より, 任意の $\nu > 0$ に対して十分大きい n で

$$\sum_{s: s \neq s^*, s \ll s^*, s \in S_n} \frac{P(s|x^n)}{P(s^*|x^n)} < \nu \quad (38)$$

である．

一方, s^* の S_n に含まれる子ノードを s^1 , その子ノードを s^2 とし以下, $s^* \ll s^1 \ll s^2 \ll \dots \in S_n$ とする．補題 3.2 の証明と同様に考えると, 仮定 3.2 より

$$\lim_{n \rightarrow \infty} \frac{P(s^{i+1}|x^n)}{P(s^i|x^n)} < O\left(\frac{1}{r}\right) \quad a.s., \quad (39)$$

である．ゆえに任意の $\mu > 0$ に対して

$$\sum_{s: s \neq s^*, s \gg s^*, s \in S_n} \frac{P(s|x^n)}{P(s^*|x^n)} < \mu \quad a.s., \quad (40)$$

となるように n を選ぶことができる．

定理 3.1 と上記の議論から任意の $\epsilon > 0$

$$\begin{aligned} & \sum_{s: s \in S_n, s \neq s^*} \frac{P(s|x^n)}{P(s^*|x^n)} \\ &= \mu + \nu < \epsilon, \quad (41) \end{aligned}$$

となる n が存在する．すなわち

$$\lim_{n \rightarrow \infty} P(s^*|x^{t-1}) = 1 \quad a.s., \quad (42)$$

が成立することがわかる．

上記は補題が成立することを意味している． □

上の定理は真のノードの事後確率が漸近的に 1 に概収束することを示している．つまり，漸近的には真のノードのみから計算した符号化確率とベイズ符号の符号化確率とは一致することがわかる．

3.2 提案アルゴリズム

前節で見たように真のノードのみを考慮すれば漸近的にベイズ符号と一致する符号化確率を得ることができる．言い換えると，事後確率最大ノードのみを考えればよいということである．しかしながら，事後確率最大ノードを探索するためにはすべてのノードを保持しておかねばならず，ベイズ符号化法と使用するノード数は変わらない．そこで，本研究では一つのポストフィクスに対して有限個のノードを用い，漸的に事後確率最大のノードを含むアルゴリズムを提案する．アルゴリズムは符号化確率の計算，事後確率の更新，文脈ノード集合の更新の三つの過程からなる．ベイズ符号化法と大きく異なるのは文脈ノード集合の更新方法であり，その文脈木のもとでは符号化確率の計算と事後確率の更新はベイズ符号化法とほぼ同様である．詳しい文脈ノード集合の更新方法は付録に示し，ここでは概要のみを述べる．なお，提案するノード集合は木ではなくグラフとなっている．それゆえ，本研究ではこのノード集合のことを文脈グラフと呼ぶ．

[準備] 提案アルゴリズムは一つのポストフィクスに対して $C(> 1)$ 個のノードのみを用いる． $M(x^{t-1})$ を $t-1$ 時点において提案アルゴリズムにより作られた文脈グラフとする． $M(x^{t-1})$ は x^{t-1} により定まる確率変数である．そして， $S_p(x^{t-1})$ を提案アルゴリズムにより作られる文脈グラフに含まれるノードと x_t のポストフィクスに対応するノードの共通集合とする．すなわち， $S_p(x^{t-1}) = \{s | M(x^{t-1}) \cap S_t\}$ であり， $|S_p(x^{t-1})| = C$ である．またきい値 $\alpha, \beta > 0$ を用意する．ここでは， x^{t-1} の条件付符号化確率 $AP_p(x_{t-1}|x^{t-2})$ まで計算済みであり， $M(x^{t-1})$ は得られているものとする．このもとで $AP_p(x_t|x^{t-1})$ の計算法，提案文脈グラフの更新法を述べる．なお，以下では $M(\cdot)$ における任意のポストフィクス $S_p(x^i)$ に対して， $\max_j s(x_{i-j}^i)$ を満たすノードを葉ノードと呼ぶことにする．更に，従来の文脈木と同様にノード間の親子関係を次のように定義する．すなわち，任意のポストフィクス $S_p(x^i)$ において $s(x_j^i), s(x_{j+1}^i) \in S_p(x^i)$ となる場合 $s(x_j^i)$ を $s(x_{j+1}^i)$ の子ノードと呼ぶことに

する．

(1) 符号化確率の計算

x^t と $M(x^{t-1})$ から $S_p(x^{t-1})$ を求め， $S_p(x^{t-1})$ における事後確率を正規化する．

$$P_p(s|x^{t-1}) = \frac{P_p(s|x^{t-1})}{\sum_{s \in S_p(x^{t-1})} P_p(s|x^{t-1})}, \quad (43)$$

正規化後の $P_p(s|x^{t-1})$ と次式を用いて符号化確率 $AP_p(x_t|x^{t-1})$ を計算する．

$$\begin{aligned} AP_p(x_t|x^{t-1}) \\ = \sum_{s \in S_p(x^{t-1})} P^s(x_t|x^{t-1}, s) P_p(s|x^{t-1}), \end{aligned} \quad (44)$$

ここで， $P^s(x_t|x^{t-1}, s)$ は式 (6) で定義される．

(2) 事後確率の更新

次式を用いて $M(x^{t-1})$ におけるノードの事後確率を更新する．

$$P_p(s|x^t) = \frac{q_p(s|x^t)}{\sum_{s \in S_p(x^{t-1})} q_p(s|x^t)}, \quad (45)$$

ここで

$$\begin{aligned} q_p(s|x^t) \\ = \begin{cases} \frac{P^s(x_t|x^{t-1}, s) P_p(s|x^{t-1})}{AP_p(x_t|x^{t-1})}, & \text{if } s \in S_p(x^{t-1}) \\ P_p(s|x^{t-1}), & \text{if } s \notin S_p(x^{t-1}) \end{cases} \end{aligned} \quad (46)$$

である．

(3) 文脈グラフの更新

[ステップ 1] $M(x^t) = M(x^{t-1})$ とする． $S_p(x^{t-1})$ における葉ノードを $s(x_j^{t-1})$ とし，その親ノードを $s(x_{j+1}^{t-1})$ とする． $\frac{P_p(s(x_{j+1}^{t-1})|x^t)}{P_p(s(x_j^{t-1})|x^t)}$ を計算し， α よりも小さければ，ノード $s(x_{j+1}^{t-1})$ を $M(x^t)$ から削除する．更に $s(x_{j+1}^{t-1})$ の子ノード集合 $\underline{M}(C, s(x_{j+1}^{t-1}))$ を $M(x^t)$ に追加する．ここで， $\underline{M}(C, s(x_{j+1}^{t-1}))$ は削除するノードより C 世代子孫にあたるすべてのノードを表す．すなわち， $|\underline{M}(C, s(x_{j+1}^{t-1}))| = 2^C$ ，である．

[ステップ 2] 追加ノードの $n(i|x^t)$ は系列を遡って数え上げる．そのため，任意の $i \in A$ に対して， $P^s(i|x^t, s)$ は従来のベイズ符号化法と同様の値となる．また，

[ステップ 3] 追加ノードの事後確率は次のようになる．

$$\begin{aligned} & P_p(s(x_{j-1}^{t-1})|x^t) \\ &= P^s(x_t|s(x_{j-1}^{t-1}))P(s(x_{j-1}^{t-1})), \end{aligned} \quad (47)$$

ここで

$$\begin{aligned} & P^s(x_t|s) \\ &= \frac{\Gamma(\beta(s))\Gamma(\beta(s|x^t) + n(0|x^t, s))}{\Gamma(\beta(x^t|s))\Gamma(\beta(s))} \\ & \cdot \frac{\Gamma(\beta(s) - \beta(s|x^t) + n(1|x^t, s))}{\Gamma(\sum_{i=0}^1 n(i|x^t, s) + \beta(s))}, \end{aligned} \quad (48)$$

である。

上記のようにして $M(x^t)$ に子孫ノードを付け加えるかどうかを決定する。一つノードを付け加えるごとに最も親ノード側にあるノードを削除するので一つのポストフィクスに対して用いるノードは常に C 個である。更に、子ノード側にノードを伸ばすのみでは真のノードより深いノードを選択してしまう場合(以降 overestimate と呼ぶ)が生じる。ステップ 2 において子孫ノードを付加しなかった場合のみ、以下のようなアルゴリズムで親ノード側に対しても判定してやることにする。

[ステップ 4] $S(x_i^{t-1})$ を $S_p(x^{t-1})$ において最も親ノード側にあるノードとする。

[ステップ 5] $\frac{P_p(s(x_{i-1}^{t-1})|x^t)}{P_p(s(x_i^{t-1})|x^t)}$ を計算し、 β より小さければノード $s(x_{i+1}^{t-1})$ を $M(x^t)$ に付け加え、 $\underline{M}(C, s(x_{i+1}^{t-1}))$ に含まれるすべてのノードを $M(x^t)$ から削除する。

[ステップ 5] 追加ノードの $n(i|x^t)$ は系列をさかのぼって数え上げる。また、追加ノードの事後確率は

$$\begin{aligned} & P_p(s(x_{i+1}^{t-1})|x^t) \\ &= \frac{P^s(x^t|s(x_{i+1}^{t-1}))P(s(x_{i+1}^{t-1}))}{AP_p(x^t)}, \end{aligned} \quad (49)$$

ここで、

$$AP_p(x^t) = \prod_{i=1}^t AP_p(x_i|x^{t-1}), \quad (50)$$

であるとする。

[注意 3.1] 符号化確率を計算する際に $M(x^t)$ における任意のポストフィクスにおいて事後確率の和が 1 である必要がある。そのため、提案アルゴリズムでは符号化確率を計算する前に式 (43) を用いて事後確率を正規化する。これは提案文脈グラフ $M(x^t)$ における任意のポストフィクス S_p に置いて $\sum_{s \in S_p} P_p(s|x^t) = 1$,

が成立しているとは限らないからである。これには次のような場合が考えられる。

ノードを文脈グラフに追加しない場合を考える。式 (45) により x^{t-1} 時点で更新された事後確率に対して $\sum_{s \in S_p(x^{t-1})} P_p(s|x^{t-1}) = 1$, が成立するが $\sum_{s \in S_p(x^t)} P_p(s|x^{t-1}) = 1$, が成立するとは限らない。例えば、 $M(x^{t-1}) = \{s(\lambda), s(0), s(1)\}$, $C = 2$ の場合を考え、 $P_p(s(\lambda)|x^{t-1}) + P_p(s(0)|x^{t-1}) = 1$, $P_p(s(\lambda)|x^{t-1}) + P_p(s(1)|x^{t-1}) = 1$, であるとする。ここで $x^{t-1} = \dots 00$ であった場合、 $S_p(x^{t-1}) = \{s(\lambda), s(0)\}$, であり、符号化確率は次のようになる。

$$\begin{aligned} AP_p(x_t|x^{t-1}) &= P^s(x_t|s(\lambda))P_p(s(\lambda)|x^{t-1}) \\ & \quad + P^s(x_t|s(0))P_p(s(0)|x^{t-1}), \end{aligned}$$

各ノードの事後確率は式 (45) より

$$P_p(s(\lambda)|x^t) = \frac{P^s(x_t|s(\lambda))P_p(s(\lambda)|x^{t-1})}{AP_p(x_t|x^{t-1})},$$

$$P_p(s(0)|x^t) = \frac{P^s(x_t|s(0))P_p(s(0)|x^{t-1})}{AP_p(x_t|x^{t-1})},$$

$$P_p(s(1)|x^t) = P_p(s(1)|x^{t-1}),$$

のように更新される。この更新により明らかに $P_p(s(\lambda)|x^t) + P_p(s(1)|x^t) = 1$, は成立していない。従来のベイズ符号化法においても同様の問題があるが、従来のベイズ符号化法においては親ノードと子ノードの事後確率の比を保存しておき、計算量を低減するアルゴリズムが提案されている [5], [6]。提案アルゴリズムにもこれらの手法が計算量の低減に有効であると考えられる。□

[注意 3.2] 追加ノードの事後確率は式 (47), 式 (49) により決定される。このアルゴリズムを用いるとステップ 1 あるいはステップ 5 の際の事後確率の比較の式において

$$\frac{P_p(s|x^t)}{P_p(s'|x^t)} = \frac{P(s|x^t)}{P(s'|x^t)}, \quad (51)$$

となることに注意されたい。□

[例 3.2] 図 2 に提案文脈グラフの例 ($C = 2$) を示す。提案文脈グラフは図の白丸ノードと点線部分の枝は保持せず、黒丸ノードのみを保持している^(注2)。現時点での提案文脈グラフが図 2 のような状態であり、

(注2): 従来のベイズ符号化法における文脈木は白丸、黒丸とすべての枝を含んだ木になる。

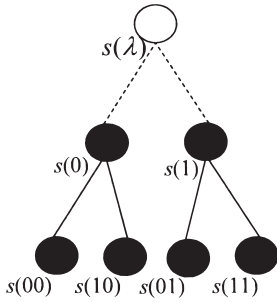


図 2 提案文脈グラフ ($C = 2$)

Fig. 2 An example of the proposed context graph.

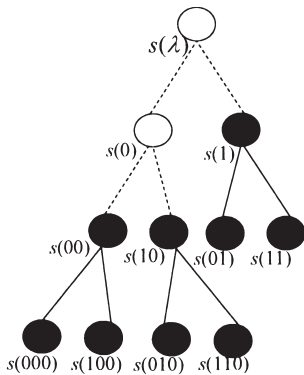


図 3 提案文脈グラフ ($C = 2$) 更新後

Fig. 3 An example of the proposed context graph.

$AP_p(x_t | \cdot 10)$ を計算したいとする．式 (44) に基づき

$$\begin{aligned} & AP_p(x_t | \cdot 10) \\ &= P^s(x_t | x^{t-1}, s(0)) P_p(s(0) | x^{t-1}) \\ & \quad + P^s(x_t | x^{t-1}, s(10)) P_p(s(10) | x^{t-1}), \quad (52) \end{aligned}$$

を計算し事後確率を更新した後， $\frac{P_p(s(0) | x^t)}{P_p(s(10) | x^t)}$ を計算， α 未満であれば，図 3 のように文脈グラフを更新する．ここで， $\underline{M}(2, s(0)) = \{s(000), s(100), s(110), s(010)\}$ である．図 2，図 3 で示すように各ポストフィクスにおいて常に $C = 2$ 個のノードが保持されている．すなわち，ここで $t + 1$ 時点のポストフィクス S_{t+1} を， $S_{t+1} = \{s(\lambda), s(0), s(10), s(110), \dots\}$ ，とすると， $S_p(x^t) = \{s(10), s(110)\}$ ，である．□

例で示したように，提案文脈グラフは各ポストフィクスにおいて C 個のノードを保持するアルゴリズムになっている．また，提案文脈グラフは根ノードをもたない場合があるので，一般的に一つの木では表すことはできない．図 3 のように従来のベイズ符号化法

における文脈木の部分木を複数含んだノードの集合になっている．

提案アルゴリズムと従来のベイズ符号化法との符号化確率の計算法の相違点は，混合をとるノード集合が異なるという点である．従来法では混合をとる集合が系列長とともに増加していったが，提案法では混合をとる集合は常に定数 C である．

4. 提案アルゴリズムの性能評価

この章では提案アルゴリズムの性能を評価する．

4.1 符号長に関する理論評価

提案アルゴリズムの性能評価に関する主定理は以下のとおり．

[定理 4.1] 仮定 3.1 を満たす任意の FSMX 情報源に対して，次を満たす N_0 が存在する． $\forall n > N_0$ において $s^* \in S_p(x^n)$ がほとんどすべての x^n で成立する．□

(定理 4.1 の証明) 提案アルゴリズムにおける $P^s(i | x^n, s)$ は系列をさかのぼって数えられていることより，任意の $s', s \in S_p(x^n)$ に対して次の等式が成立する．

$$\begin{aligned} \frac{P_p(s' | x^n)}{P_p(s | x^n)} &= \frac{P(x^n | s') P(s')}{P(x^n | s) P(s)} \\ &= \frac{\sum_{\{m | s' \in S(m)\}} P(x^n | m) P(m)}{\sum_{\{m | s \in S(m)\}} P(x^n | m) P(m)}. \quad (53) \end{aligned}$$

2 番目の等式は $P(s | x^n)$ の定義より成立する．

ここで，定理 3.1 の証明と同様に考えると，式 (53) の右辺は 0 に概収束する．ゆえに式 (53) より， $s' \ll s \ll s^*$ であるような任意の $s' \ll s$ に対して一様に次が成立する．すなわち， $s', s \in S_p$ を満たすほとんどすべての x^n に対して

$$\lim_{n \rightarrow \infty} \frac{P_p(s' | x^n)}{P_p(s | x^n)} < \alpha \quad a.s., \quad (54)$$

が成立する．同様に式 (53) と補題 4.1 より， $s^* \ll s \ll s'$ であるような任意の $s' \ll s$ に対して一様に次が成立する．すなわち， $s', s \in S_p$ を満たすほとんどすべての x^n に対して

$$\lim_{n \rightarrow \infty} \frac{P_p(s' | x^n)}{P_p(s | x^n)} < \beta \quad a.s., \quad (55)$$

が成立する．これより， α, β が 0 より大きければ提案アルゴリズムは確率 1 で s^* を含むことがわかる．ゆ

えに定理は証明された。 □

更に、次の補題が成立する。

[補題 4.1] $s, s^* \in S_n, \forall s \neq s^*$ を満たす任意の s, s^* に対して仮定 3.1 のもとで、一様に次が成立する。すなわち、 $s^*, s \in S_p$ を満たすほとんどすべての x^n に対して

$$\lim_{n \rightarrow \infty} \left| \frac{P_p(s|x^n)}{P_p(s^*|x^n)} \right| = 0 \quad a.s., \quad (56)$$

が成立する。

(証明) 定理 4.1 の証明と同様に

$$\begin{aligned} \frac{P_p(s|x^n)}{P_p(s^*|x^n)} &= \frac{P(x^n|s)P(s)}{P(x^n|s^*)P(s^*)} \\ &= \frac{\sum_{\{m|s \in S(m)\}} P(x^n|m)P(m)}{\sum_{\{m|s^* \in S(m)\}} P(x^n|m)P(m)}, \end{aligned} \quad (57)$$

ここで、 $s^* \in S(m^*)$ であることを考えると、式 (57) は、

$$\frac{P_p(s|x^n)}{P_p(s^*|x^n)} \leq \frac{\sum_{\{m|s \in S(m)\}} P(x^n|m)P(m)}{P(x^n|m^*)P(m^*)} \quad (58)$$

が成立する。ここで定理 3.1 と同様に考えると上式右辺は 0 に収束する。ゆえに補題が証明された。 □

定理 4.1 と補題 4.1 より提案アルゴリズムを用いたときの符号化確率が漸近的にベイズ符号化法の符号化確率と一致することがわかる。すなわち次の定理が成り立つ。

[定理 4.2] 仮定 3.1, 3.2 のもとで

$$\lim_{t \rightarrow \infty} -\log \frac{AP_p(x_t|x^{t-1})}{AP(x_t|x^{t-1})} = 0 \quad a.s., \quad (59)$$

が成り立つ。 □

(定理 4.2 の証明) 定理 4.1 より、提案アルゴリズムを用いたとき S_p は確率 1 で s^* を含むことがわかる。このことと補題 4.1、 $|S_p| = C$ であることより、

$$\lim_{t \rightarrow \infty} P_p(s^*|x^{t-1}) = 1 \quad a.s., \quad (60)$$

であることがわかる。ゆえに、式 (44) より

$$\begin{aligned} \lim_{t \rightarrow \infty} \left| AP_p(x_t|x^t - 1) - P^s(x_t|s^*(x^{t-1}), x^{t-1}) \right| \\ = 0 \quad a.s., \end{aligned} \quad (61)$$

を得る。これより定理は証明された。 □

上記の定理は提案アルゴリズムのシンボル当りの符号長が漸近的にベイズ符号化法のシンボル当りの符号長と一致することを示している。

また漸近的には、しきい値 α, β は $0 < \alpha, \beta < \infty$ の間の任意の値でよいこと、 $C > 1$ の個数も任意でよいことが定理の証明からわかる。

4.2 メモリ容量・計算量に関する評価

提案アルゴリズムの部分で述べたように提案文脈グラフは各ポストフィクスにおいて C 個のノードのみを保持している。また、前節で示したように十分大きい n のもとで、各ポストフィクスにおいて提案文脈グラフは真のノードを確率 1 で含む。ゆえに FSMX 情報源 (m^*, θ^{m^*}) に対して提案アルゴリズムはたかだか $C|S(m^*)| + \epsilon$ 個のノードを保持しておけばよい。ここで、 ϵ は overestimate してしまった際に余分に必要になるノードの個数を表している。

一方、従来のベイズ符号化法は逐次的に文脈木を生成していくため、必要なノードの個数は系列長 n の指数オーダーとなる。

計算量の面を考えると、従来のベイズ符号化法は混合をとる集合が増加するために計算量も系列長 n とともに増加していく。その量は $O(n)$ である。一方、提案アルゴリズムはノードを付け加えるかどうかの判定の分の計算量が $O(1)$ 、符号化確率 $AP_p(x_n|x^{n-1})$ を計算するための計算量は C のみに依存して、 n には依存しないので、 $O(1)$ である。更にノードを付け加える際の $P^s(x_n|s, x^{n-1})$ を計算するための計算量は過去の系列 x^{n-1} を一度参照し、数え上げるので $O(n)$ である。これらをすべて加えると提案アルゴリズムに必要な計算量はたかだか $O(n)$ であることがわかる。

つまり、提案アルゴリズムは従来法に比べ、オーダーで考えると計算量は同等でメモリ容量が少なくて済むことがわかる。

また、FSMX モデルを選択するという観点からいうと、提案アルゴリズムは Rissanen の提唱した記述長最小基準 [9] に基づく符号 (以下 MDL 符号) とも関係が深い。ここで、MDL 符号もモデルを探索する際にすべてのノードを保持しておく必要があり、ベイズ符号と同等のメモリ量が必要になる。そのため、本研究と同様の問題設定に対してはメモリ量が系列長とともに指数的に増加してしまうという問題があることに注意されたい。

4.3 数値実験による評価

4.3.1 実験条件

前節で示したように漸的には提案アルゴリズムを用いた際の符号長はベイズ符号の符号長と一致する．本節では有限時点での提案アルゴリズムの性能を数値実験により評価する．実験には最大深さ 9 の binary-FSMX 情報源 (状態数 28) を用い, 50 系列長に対する 1 文字当りの符号長の平均をとった．実験 1 では提案アルゴリズムの符号長を評価する実験を行った．この結果を図 4 に示す．Bayes はベイズ符号化法, MDL は MDL 符号, $C=3$ は $\alpha = \beta = 0.3$, $C=3$ の場合を表している．

実験 2 では α と β の違いに提案アルゴリズムの性能を見た．使用した情報源は実験 1 と同様である．この情報源に対して 1000 時点ごとに真のノード 28 個のうちいくつか提案文脈グラフに含まれるかを調べた．条件は $\alpha = \beta = 0.1$, $\alpha = \beta = 0.3$, $\alpha = \beta = 0.5$ である．この結果を図 5 に示す．

4.3.2 実験結果に関する考察

実験結果 1 より, 提案アルゴリズムを用いた場合の 1 文字当りの符号長が漸的にベイズ符号の 1 文字当りの符号長に近づくことが確認できた．また, 提案手法は MDL 符号とベイズ符号の中間の性能を示している．実際には C や α, β の値を変更していくつか実験を行ったが, ベイズ符号に近い性能を示すか MDL 符号により近い性能を示すかの違いはあったが, すべて同様の結果を示した．これは, 提案アルゴリズムが使用するモデルの点からは MDL 符号とベイズ符号の中間に位置することによって考えられる．また, ベイズ符号, MDL 符号を実行するために必要なメモリ量は入力系列長に対し指数的に増加していくことを考えると, この結果は非常に意味がある結果だと考えられる．

実験結果 2 に対しては, 理論上は α も β も正の値であれば漸的に性能は変わらない．しかしながら, その違いは符号長の収束速度に関係していると考えられる．結果より, 系列数が短い時点では α, β が大きい値の方がより多く真のノードを含んでいることがわかる．しかし, α, β が大きい場合, 系列長が長くなっても個数があまり増えていない．これは α, β が大きい場合, overestimate してしまう可能性が非常に大きくなるためだと考えられる． α, β の値に関しては符号長によって α, β の値を変化させることも考えると非常に多くのバリエーションが考えられるが, 本研究ではこれについては言及しない．

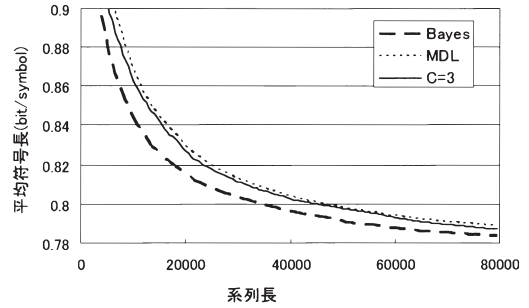


図 4 提案アルゴリズムとベイズ符号の平均符号長
Fig.4 Codelength of the proposed algorithm and Bayes code.

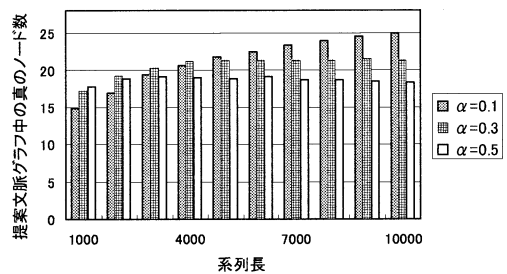


図 5 提案文脈グラフに含まれる真のノード数
Fig.5 Number of true nodes included in the proposed context graph.

5. むすび

本研究ではメモリ容量を低減したベイズ符号化法の近似計算アルゴリズムを提案した．更に, 提案アルゴリズムのシンボル当りの符号長がベイズ符号化法の符号長と一致することを示した．ベイズ符号化法はベイズ基準のもとで冗長度を最小にする符号であるが, すべての候補ノードを用い符号化確率を求めるため, 実用化にはメモリ容量, 計算量の点を低減する必要がある．そのため, 本研究のようにメモリ容量を低減した近似的なベイズ符号化アルゴリズムは実用化のために意味があると思われる．

謝辞 本研究に関して貴重な御意見を頂きました神奈川工科大学新家稔央氏, 武蔵工業大学後藤正幸氏, 並びに早稲田大学松嶋研究室, 平澤研究室の各位に深く感謝致します．本研究の一部は, 文部省科学研究費基盤 (C)(No.12650400), 早稲田大学特定課題研究助成費 (2001A-570) の援助による．

文 献

- [1] L.D. Davison, "Universal noiseless coding," IEEE

- Trans. Inf. Theory, vol.19, no.6, pp.783–795, 1973.
- [2] J. Ziv and H. Lempel, “A universal algorithm for sequential data compression,” IEEE Trans. Inf. Theory, vol.23, no.3, pp.337–343, 1977.
- [3] T. Matsushima, H. Inazumi, and S. Hirasawa, “A class of distortionless codes designed by bayes decision theory,” IEEE Trans. Inf. Theory, vol.37, no.5, pp.1288–1293, 1991.
- [4] F.M.J. Willems, Y.M. Shtarkov, and T.J. Tjalkens, “The context tree weighting method: Basic properties,” IEEE Trans. Inf. Theory, vol.41, no.3, pp.653–663, 1995.
- [5] T. Matsushima and S.Hirasawa, “A bayes coding using context tree,” Proc. Int. Symp. on. Inf. Theory, p.386, 1994.
- [6] T. Kawabata and F.M.J. Willems, “A context tree weighting algorithm with an incremental context set,” IEICE Trans. Fundamentals, vol.E83-A, no.10, pp.1898–1903, Oct. 2000.
- [7] T. Matsushima and S. Hirasawa, “A Bayes coding Algorithm for Markov models,” IEICE Technical Report, IT95-1, 1995.
- [8] J. Ziv and H. Lempel, “Compression of individual sequences via variable-rate coding,” IEEE Trans. Inf. Theory, vol.24, no.5, pp.530–536, 1978.
- [9] J. Rissanen, “Modeling by shortest data description,” Automatica, vol.46, pp.465–471, 1978.
- [10] W. Feller, 確率論とその応用, 紀伊国屋書店, 1960.
- [11] 韓 太舜, 小林欣吾, 情報と符号化の数理, 岩波書店, 1994.
- [12] B.S. Clarke and A.R. Barron, “Information theoretic asymptotics of Bayes methods,” IEEE Trans. Inf. Theory, vol.36, no.3, pp.453–471, 1990.
- [13] M. Gotoh, T. Matsushima, and S. Hirasawa, “A generalization of B.S. Clarke and A.R. Barron’s asymptotics of Bayes codes for FSMX sources” IEICE Trans. Fundamentals, vol.E81-A, no.10, Oct. 1998.

付 録

1. 提案文脈木更新アルゴリズム

begin

$$M(x^t) = M(x^{t-1})$$

$$j := \arg \min_j s(x_j^{t-1}) \in M(x^t);$$

$$nc = C;$$

/* A decision to add child node */

if $\frac{P_p(s(x_{j+1}^{t-1})|x^t)}{P_p(s(x_j^{t-1})|x^t)} < \alpha$ **then**

add $\underline{M}(C, s(x_{j+C-1}^{t-1}))$ to $M(x^t)$

delete $s(x_{j+C-1}^{t-1})$ from $M(x^t)$;

else

while $nc \geq 0$ **do**

$$j := j - 1;$$

$$nc := nc - 1;$$

end-while

/* A decision to add ancestor node */

if $\frac{P_p(s(x_j^{t-1})|x^t)}{P_p(s(x_{j+1}^{t-1})|x^t)} < \beta$ **then**

add $s(x_{j+1}^{t-1})$ to $M(x^t)$;

$$nc = C;$$

delete all s in $\underline{M}(C, s(x_{j+1}^{t-1}))$ from $M(x^t)$;

end

2. 補題の証明

任意の $m_{s'}$ について,

$$\lim_{n \rightarrow \infty} \frac{P(x^n | m_{s'})}{P(x^n | m_s)} = 0 \quad a.s., \quad (\text{A.1})$$

となる m_s が存在することを示す. 性質 3.2 より,

$$\begin{aligned} P(\hat{\theta}^m | m, x^n) \\ = \left(\frac{n}{2\pi} \right)^{\frac{k_m}{2}} \sqrt{\det I(\hat{\theta} | m) + o(n^{\frac{k_m}{2}})}, \end{aligned} \quad (\text{A.2})$$

である. また,

$$P(x^n | m) = \frac{P(x^n | m, \hat{\theta}^m) P(\hat{\theta}^m | m)}{P(\hat{\theta}^m | m, x^n)}, \quad (\text{A.3})$$

であるので, 任意の m_s と $m_{s'}$ に対して次の式が成り立つ.

$$\begin{aligned} \log \frac{P(x^n | m_{s'})}{P(x^n | m_s)} \\ = \log \frac{P(x^n | m_{s'}, \hat{\theta}^{m_{s'}})}{P(x^n | m_s, \hat{\theta}^{m_s})} + \frac{k_{m_s} - k_{m_{s'}}}{2} \log \frac{n}{2\pi} \\ - \log \frac{\sqrt{\det I(\hat{\theta}^{*m_{s'}} | m_{s'})}}{\sqrt{\det I(\hat{\theta}^{*m_s} | m_s)}} + \log \frac{P(\hat{\theta}^{*m_{s'}} | m_{s'})}{P(\hat{\theta}^{*m_s} | m_s)} \\ + o(1). \end{aligned} \quad (\text{A.4})$$

ここで, $\log \frac{\sqrt{\det I(\hat{\theta}^{*m_{s'}} | m_{s'})}}{\sqrt{\det I(\hat{\theta}^{*m_s} | m_s)}}$, $\log \frac{P(\hat{\theta}^{*m_{s'}} | m_{s'})}{P(\hat{\theta}^{*m_s} | m_s)}$ は n に無関係な定数であることより以下を得る.

$$\begin{aligned} \log \frac{P(x^n | m_{s'})}{P(x^n | m_s)} \\ = \log \frac{P(x^n | m_{s'}, \hat{\theta}^{m_{s'}})}{P(x^n | m_s, \hat{\theta}^{m_s})} \\ + \frac{k_{m_s} - k_{m_{s'}}}{2} \log \frac{n}{2\pi} + O(1). \end{aligned} \quad (\text{A.5})$$

次に $-\log P(x^n | m_s, \hat{\theta}^{m_s})$ を評価する.

まず, $-\log P(x^n|m_s, \hat{\theta}^{m_s})$ をテイラー展開すると,

$$\begin{aligned}
 & -\log P(x^n|m_s, \theta^{*m_s}) \\
 &= -\log P(x^n|m_s, \hat{\theta}^{m_s}) \\
 & \quad - \left. \frac{\partial \log P(x^n|m_s, \theta^{m_s})}{\partial \theta^{m_s}} \right|_{\theta^{m_s}=\hat{\theta}^{m_s}} (\theta^{*m_s} - \hat{\theta}^{m_s}) \\
 & \quad + \frac{1}{2} (\theta^{*m_s} - \hat{\theta}^{m_s})^T \\
 & \quad \left. \frac{\partial^2 \log P(x^n|m_s, \theta^{m_s})}{\partial \theta^{m_s T} \partial \theta^{m_s}} \right|_{\theta^{m_s}=\hat{\theta}^{m_s}} (\theta^{*m_s} - \hat{\theta}^{m_s}) \\
 & \quad + o((\theta^{*m_s} - \hat{\theta}^{m_s})^T \\
 & \quad \left. \frac{\partial^2 \log P(x^n|m_s, \theta^{m_s})}{\partial \theta^{m_s T} \partial \theta^{m_s}} \right|_{\theta^{m_s}=\hat{\theta}^{m_s}} (\theta^{*m_s} - \hat{\theta}^{m_s})). \tag{A.6}
 \end{aligned}$$

ここで, $\left. \frac{\partial \log P(x^n|m_s, \theta^{m_s})}{\partial \theta^{m_s}} \right|_{\theta^{m_s}=\hat{\theta}^{m_s}} = 0$ である. また, 大数の強法則より次の式が概収束の意味で成り立つ.

$$\begin{aligned}
 & -\frac{1}{n} \left. \frac{\partial^2 \log P(x^n|m_s, \theta^{m_s})}{\partial \theta^{m_s T} \partial \theta^{m_s}} \right|_{\theta^{m_s}=\hat{\theta}^{m_s}} \\
 &= I(\hat{\theta}^{m_s}|m_s) + o(1). \tag{A.7}
 \end{aligned}$$

つまり, $\frac{\partial^2 \log P(x^n|m_s, \theta^{m_s})}{\partial \theta^{m_s T} \partial \theta^{m_s}}$ の部分は, Fisher 情報量の n 倍に概収束する. 次に, $(\theta^{*m_s} - \hat{\theta}^{m_s})$ の部分を考える. 系 3.1 より,

$$\sqrt{n} \|\hat{\theta}^{m_s} - \theta^{*m_s}\| = O\left((\log \log n)^{\frac{1}{2}}\right), \tag{A.8}$$

であるので, 結局以下を得る.

$$\begin{aligned}
 & \frac{n}{2} (\theta^{*m_s} - \hat{\theta}^{m_s})^T I(\hat{\theta}^{m_s}|m_s) (\theta^{*m_s} - \hat{\theta}^{m_s}) \\
 &= O(\log \log n) \text{ a.s.} \tag{A.9}
 \end{aligned}$$

式 (A.9) を式 (A.6) に代入すると,

$$\begin{aligned}
 & \log P(x^n|m_s, \hat{\theta}^{m_s}) \\
 &= \log P(x^n|m_s, \theta^{*m_s}) + O(\log \log n) \text{ a.s.} \tag{A.10}
 \end{aligned}$$

である. 同様のことが $P(m_{s'}|x^n)$ にも成立するので結局, 式 (17), (A.5), (A.10) から,

$$\begin{aligned}
 & \log \frac{P(x^n|m_{s'})}{P(x^n|m_s)} \\
 &= \log \frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_s, \theta^{*m_s})}
 \end{aligned}$$

$$\begin{aligned}
 & + O(\log \log n) + \frac{k_{m_s} - k_{m_{s'}}}{2} \log \frac{n}{2\pi} \\
 &= -nD(p_{m_{s'}}^*; p_{m_s}^*) + O(\log \log n) \\
 & \quad + \frac{k_{m_s} - k_{m_{s'}}}{2} \log \frac{n}{2\pi} \text{ a.s.}, \tag{A.11}
 \end{aligned}$$

である.

ここで, $s^* \ll s \ll s'$ の場合を考える. 明らかに任意の $m_{s'}$ に対して, 以下を満たす m_s が存在する.

$$P(x^n|m_s, \theta^{*m_s}) = P(x^n|m_{s'}, \theta^{*m_{s'}}). \tag{A.12}$$

上式を満たす m_s と $m_{s'}$ を考えると, $k_{m_{s'}} > k_{m_s}$ より,

$$\begin{aligned}
 & \log \frac{P(x^n|m_{s'})}{P(x^n|m_s)} \\
 &= \log \frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_s, \theta^{*m_s})} + \frac{k_{m_s} - k_{m_{s'}}}{2} \log \frac{n}{2\pi} \\
 & \quad + O(\log \log n) \\
 &= 0 + \frac{k_{m_s} - k_{m_{s'}}}{2} \log \frac{n}{2\pi} + O(\log \log n) \\
 & \rightarrow -\infty \text{ a.s.} \tag{A.13}
 \end{aligned}$$

となる.

一方, $s' \ll s \ll s^*$ の場合, $k_{m_{s'}} < k_{m_s}$ である. ここで,

$$\begin{aligned}
 & \frac{1}{n} \log \frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_s, \theta^{*m_s})} \\
 &= \frac{1}{n} \log \frac{\frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_s, \theta^{*m_s})}}{\frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_{s'}, \theta^{*m_{s'}})}}, \\
 &= \frac{1}{n} \log \frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_s, \theta^{*m_s})} \\
 & \quad - \frac{1}{n} \log \frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_{s'}, \theta^{*m_{s'}})}. \tag{A.14}
 \end{aligned}$$

が成立する. また仮定 3.1 より,

$$D(p_{m_{s'}}^*; p_{m_s}^*) - D(p_{m_s}^*; p_{m_{s'}}^*) > 0, \tag{A.15}$$

であるので, 式 (17) より, 任意の $m_{s'}$ に対して

$$\begin{aligned}
 & \log \frac{P(x^n|m_{s'}, \theta^{*m_{s'}})}{P(x^n|m_s, \theta^{*m_s})} \\
 &= n \left(D(p_{m_{s'}}^*; p_{m_s}^*) - D(p_{m_s}^*; p_{m_{s'}}^*) \right) + o(n) \\
 & \rightarrow -\infty \text{ a.s.}, \tag{A.16}
 \end{aligned}$$

となる m_s が存在することがわかる.

以上の議論より,

$$\begin{aligned}
 & \log \frac{P(x^n | m_{s'})}{P(x^n | m_s)} \\
 &= \log \frac{P(x^n | m_{s'}, \theta^{*m_{s'}})}{P(x^n | m_s, \theta^{*m_s})} + \frac{k_{m_{s'}} - k_{m_s}}{2} \log \frac{n}{2\pi} \\
 & \quad + O(\log \log n) \\
 &= -O(n) + \frac{k_{m_{s'}} - k_{m_s}}{2} \log \frac{n}{2\pi} \\
 & \quad + O(\log \log n) \\
 & \rightarrow -\infty \text{ a.s.} \tag{A.17}
 \end{aligned}$$

を得る. $-O(n)$ は負の数を表す.

上式より任意の $m_{s'}$ に対して次式を満たす m_s が存在することがわかる.

$$\lim_{n \rightarrow \infty} \frac{P(x^n | m_{s'})}{P(x^n | m_s)} = 0 \text{ a.s.}, \tag{A.18}$$

ゆえに補題は証明された. \square

(平成 13 年 4 月 13 日受付, 8 月 24 日再受付,
9 月 20 日最終原稿受付)



平澤 茂一 (正員)

昭 36 早大・理工・数学卒. 昭 38 同電気通信卒. 同年三菱電機(株)入社. 昭 56 早大・理工・工業経営学科(現在, 経営システム工学科)教授, 現在に至る. 情報理論とその応用, データ伝送方式, 並びに計算機応用システムの開発などの研究に従事. 工博. 昭 54 UCLA 計算機科学科客員研究員. 昭 60 ハンガリー科学アカデミー, 昭 61 伊トリエステ大学客員研究員. 平 5 本会小林記念特別賞, 業績賞受賞. IEEE Fellow. 情報理論とその応用学会, 人工知能学会, 情報処理学会, OR 学会, 日本経営工学会等各会員.



野村 亮 (正員)

平 8 早大・理工・工業経営卒. 平 10 同大学院修士課程了. 平 11 同大学院博士後期課程入学. 平 12 同大助手. 情報源符号化に関する研究に従事. IEEE, 情報理論とその応用学会, 日本経営工学会等各会員.



松嶋 敏泰 (正員)

昭 53 早大・理工・工業経営卒. 昭 55 同大学院修士課程了. 同年, 日本電気(株)入社. 昭 61 早大・理工学研究科・博士後期課程入学. 平 1 横浜商科大学講師. 平 3 同大助教授. 平 4 早大・理工学部・工業経営学科(現在, 経営システム工学科)助教授, 平 9 同大教授, 現在に至る. 知識情報処理及び情報理論とその応用に関する研究に従事. 工博. 平 13 ハワイ大学客員研究員. IEEE, 情報理論とその応用学会, 人工知能学会, 情報処理学会, OR 学会, 日本経営工学会等各会員.