PAPER

Universal Variable-to-Fixed Length Codes Achieving Optimum Large Deviations Performance for Empirical Compression Ratio

Tomohiko UYEMATSU†, *Member*

SUMMARY This paper clarifies two variable-to-fixed length codes which achieve optimum large deviations performance of empirical compression ratio. One is Lempel-Ziv code with fixed number of phrases, and the other is an arithmetic code with fixed codeword length. It is shown that Lempel-Ziv code is asymptotically optimum in the above sense, for the class of finite-alphabet and finite-state sources, and that the arithmetic code is asymptotically optimum for the class of finite-alphabet unifilar sources.
key words: source coding, variable-to-fixed length code, empirical compression rate, finite state source

1. Introduction

Comparisons between lossless variable-to-fixed (V-F) length codes and fixed-to-variable (F-V) length codes have been done by many researchers. Especially, Ziv [1] has shown that for Markovian sources with long memory there exists a V-F length code that provides a better compression ratio than any F-V length code with the same number of codewords. Tjalkens and Willems [2] have proved similar results for universal coding of binary memoryless sources. Finally, Merhav and Neuhoff [3] have shown that for unifilar sources the best V-F length code provides a better large deviations performance than any F-V length code with the same number of codewords. Especially, they have considered a random variable, referred to as the empirical compression ratio (ECR), which is defined as the length in bits of the encoder output word divided by the length in bits of the input word. It has been shown that the exponential decay rate of the probability that the ECR exceeds R, for the best V-F length code, is $1/R$ times faster than that of the best F-V length code with the same number of codewords.

In this paper, we clarify that for finite-state sources Lempel-Ziv (LZ) code [4] with fixed number of phrases achieves the optimum large deviations performance, more precisely, asymptotically minimizes the probability that the ECR exceeds R among V-F length code with the same codelength. Further, for unifilar sources, an adaptive version of arithmetic codes [5]–[8] with fixed code length also achieves the optimum large deviations performance.

Of related works, we mention the following two results. Teuhola and Raita [9] proposed a V-F code using an arithmetic coding, which was very similar to our proposed code. However, they were interested in the practical performance of the coding and did not investigate its asymptotic performance. Visweswariah and Kulkarni [10] have proposed an universal V-F code for the class of Markovian sources which is asymptotically optimum. However, their criterion of optimality is not ECR but redundancy.

Manuscript received Oct. ??, 1998.
Manuscript revised April ??, 1999.
†The author is with the Department of Electrical and Electronic Eng., Tokyo Institute of Technology, Tokyo 152-8552 Japan.

2. Optimum Code for Finite-State Sources

Let $\boldsymbol{x} = x_1 x_2 \cdots x_i \cdots x_n$ be a sequence of observable random variables taking values in a finite alphabet $\mathcal{X}$. Similarly, let $\boldsymbol{s} = s_1 s_2 \cdots s_i \cdots s_n$ be another sequence of random variables, called states, which take values in another finite set $\mathcal{S}$. A probabilistic source P is called finite-state (with $|\mathcal{S}|$ states) if

$$P(\boldsymbol{x}, \boldsymbol{s}) = \prod_{i=1}^{n} P(x_i, s_i | s_{i-1}), \tag{1}$$

where $P(x_i, s_i|s_{i-1})$ is the joint probability of a letter x_i and a state s_i given the previous state s_{i-1}, $P(\boldsymbol{x}, \boldsymbol{s})$ is the joint probability of $\boldsymbol{x}$ and $\boldsymbol{s}$, and $s_0 \in \mathcal{S}$ is a fixed initial state. The class of finite-state sources with no more than S states will be denoted by $\mathcal{P}_S$

Let $\mathcal{X}^*$ denote the set of all words over alphabet $\mathcal{X}$. A set of words $Z \subset \mathcal{X}^*$ is said to be prefix set. We require it to be *proper* and *complete*. Properness of the prefix set implies that no word in Z is a *prefix* of any other word in Z. Completeness guarantees that each infinite sequence $x_1 x_2 \cdots$ has one and only one prefix that belongs to Z. Any prefix set Z can be represented by a complete $|\mathcal{X}|$-ary tree $T(Z)$, where $T(Z)$ has Z as the set of terminal nodes (cf. e.g. [11]).

A V-F length code $f : \mathcal{X}^* \to \{0,1\}^n$ can be characterized by a prefix set $\{X_1, X_2, \cdots, X_N\} \subset \mathcal{X}^*$ with $N(\leq 2^n)$, which can be represented by a complete $|\mathcal{X}|$-ary tree T_n with N terminal nodes, and the one-to-one mapping from each prefix (or terminal node) X_i ($i = 1, 2, \cdots, N$) to binary codeword of length n bits. From the definition, it is obvious that every infinite word X has one and only one prefix $X_i \in T_n$. In what follows, for infinitely long sequence X, we shall adopt the notation $\ell(X)$ as the length of the prefix of X in T_n, namely $\ell(X) \triangleq \ell(X_i)$ where X_i is the prefix of X in T_n. Then, the empirical compression ratio (ECR) associated with a V-F length code $f : \mathcal{X}^* \to \{0,1\}^n$ and an infinite source sequence X is defined as

$$\rho_f(X) \triangleq \frac{n}{\ell(X)}.$$

Given a finite-state source P and a constant $R > 0$, we consider a problem to find a V-F length code $f : \mathcal{X}^* \to \{0,1\}^n$ that minimizes the probability that the ECR exceeds R, defined as

$$\Pr\{\rho_f(X) > R\} \triangleq \sum_{\boldsymbol{x} \in T_n : \ell(\boldsymbol{x}) > nR} P(\boldsymbol{x}),$$

where $P(\boldsymbol{x}) = \sum_{\boldsymbol{s} \in \mathcal{S}^n} P(\boldsymbol{x}, \boldsymbol{s})$ and $P(\boldsymbol{x}, \boldsymbol{s})$ is in (1). It is assumed that $H(P) < R < \log_2 |\mathcal{X}|$, where $H(P)$ is the Shannon entropy, which for a stationary source is given by

$$H(P) \triangleq -\lim_{n\to\infty} \frac{1}{n} \sum_{\boldsymbol{x} \in \mathcal{X}^n} P(\boldsymbol{x}) \log P(\boldsymbol{x}).$$

Now, consider following V-F code which is a V-F version of the Lempel-Ziv code [4].

LZ code with fixed codelength $f_{LZ} : \mathcal{X}^* \to \{0,1\}^{n_m}$

1) (Initialize) Let $m(> 1)$ be an integer which specifies the codelength. Let $x_1 x_2 \cdots x_i \cdots$ be an output sequence of the source. Then, let $w(1) \leftarrow x_1$, $j \leftarrow 1$ and $i_j \leftarrow 1$. Output x_1 in $\lceil \log_2 |\mathcal{X}| \rceil$-bit †.
2) (Parsing) Suppose $w(1) \cdots w(j) = x_1 \cdots x_{i_j}$.
   - (i) If $x_{i_j+1} \notin \{w(1), \cdots, w(j)\}$ then $w(j+1) \leftarrow x_{i_j+1}$ and $p \leftarrow 0$.
   - (ii) Otherwise, $w(j+1) = x_{i_j+1} \cdots x_{k+1}$, where k is the least integer greater than i_j such that $x_{i_j+1} \cdots x_k = w(p)$ for some $1 \leqq p \leqq j$ and $x_{i_j+1} \cdots x_{k+1} \notin \{w(1), \cdots, w(j)\}$.
3) Output a pair (p, x_{k+1}) in $\lceil \log_2 (j+1)|\mathcal{X}| \rceil$-bit. If $j < m$ then $j \leftarrow j+1$, $i_j \leftarrow k+1$ and go to 2). Otherwise, terminate the algorithm. □

The codelength of this code is given by $n_m \triangleq \sum_{j=1}^{m} \lceil \log_2 j|\mathcal{X}| \rceil$. The following theorem establishes the asymptotic optimality of the LZ code in the sense of minimum probability of the ECR exceeding R.

Theorem 1: For every finite-state source $P \in \mathcal{P}_S$, any V-F length code $f : \mathcal{X}^* \to \{0,1\}^{n_m}$ with ECR $\rho_f(X)$, every $R \in (H(P), \log_2 |\mathcal{X}|)$, and all large m,

$$\begin{aligned} &\Pr\{\rho_{LZ}(X) > R + \epsilon(\lceil n_m/R \rceil)\} \\ &\leqq (1 + \eta(\lceil n_m/R \rceil)) \Pr\{\rho_f(X) > R\}, \end{aligned}$$

where $\rho_{LZ}(X)$ is the ECR of LZ code with a fixed codelength n_m, $\eta(n) = n^2 2^{-n/\sqrt{\log n}}$ and $\epsilon(n) = O(1/\sqrt{\log n})$ is a positive sequence depending on $\mathcal{X}$ and $\mathcal{S}$.

Theorem 1 states that the LZ code achieves the optimum ECR performance, i.e. the best large deviations performance.

Before the proof of Theorem 1, we introduce a proposition obtained by Merhav [12] which shows the dual result of ours. That is, LZ code yields the shortest length, uniformly for every sufficiently long sequence $\boldsymbol{x}$, among all information lossless F-V codes for any finite-state source.

Proposition [12, Theorem 1]: For any F-V length code, let $L_n(\boldsymbol{x})$ denote the codelength for $\boldsymbol{x}(\in \mathcal{X}^n)$. Then, for every $B \in (H(P), \log_2 |\mathcal{X}|)$, every finite-state source $P \in \mathcal{P}_S$, and all large n,

$$\begin{aligned} &\Pr\{\boldsymbol{x} \in \mathcal{X}^n : n^{-1} U_{LZ}(\boldsymbol{x}) > B + \epsilon(n)\} \\ &\leqq (1 + \eta(n)) \Pr\{\boldsymbol{x} \in \mathcal{X}^n : n^{-1} L_n(\boldsymbol{x}) > B\}, \end{aligned}$$

where $U_{LZ}(\boldsymbol{x})$ denote the codelength of LZ code (when used as F-V code) for a sequence $\boldsymbol{x} \in \mathcal{X}^n$, while $\eta(n)$ and $\epsilon(n)$ are given in Theorem 1.

Proof of Theorem 1: It is easy to see

$$\begin{aligned} &\Pr\{\rho_{LZ}(X) > R\} \\ &= \Pr\{\boldsymbol{x} \in \mathcal{X}^{\lceil n_m/R \rceil} : U_{LZ}(\boldsymbol{x}) > R\lceil n_m/R \rceil\} \\ &= \Pr\{\boldsymbol{x} \in \mathcal{X}^{\lceil n_m/R \rceil} : (\lceil n_m/R \rceil)^{-1} U_{LZ}(\boldsymbol{x}) > R\}. \end{aligned} \quad (2)$$

On the other hand, let us consider any V-F length code $f : \mathcal{X}^* \to \{0,1\}^{n_m}$ with a complete prefix tree T_{n_m}. By using the method introduced by Merhav and Neuhoff [3, Proof of Theorem 2] (see also Appendix), we can construct a F-V code $\tilde{C}_{n_m}$ with block length $\lceil n_m/R \rceil$ from a given V-F code f such that

$$\begin{aligned} &\Pr\{\boldsymbol{x} \in T_n : \ell(\boldsymbol{x}) < n_m/R\} \\ &= \Pr\{\boldsymbol{x} \in \mathcal{X}^{\lceil n_m/R \rceil} : \tilde{L}(\boldsymbol{x}) > n_m\} \end{aligned}$$

where $\tilde{L}(\boldsymbol{x})$ denotes a length of codeword in $\tilde{C}_{n_m}$ corresponding to $\boldsymbol{x}(\in \mathcal{X}^{\lceil n_m/R \rceil})$. Hence,

$$\begin{aligned} &\Pr\{\rho_f(X) > R\} \\ &= \Pr\{\boldsymbol{x} \in T_{n_m} : \ell(\boldsymbol{x}) < n_m/R\} \\ &= \Pr\{\boldsymbol{x} \in \mathcal{X}^{\lceil n_m/R \rceil} : \tilde{L}(\boldsymbol{x}) > n_m\} \\ &\geqq \Pr\{\boldsymbol{x} \in \mathcal{X}^{\lceil n_m/R \rceil} : (\lceil n_m/R \rceil)^{-1} \tilde{L}(\boldsymbol{x}) > R\} \\ &\geqq \frac{1}{1 + \eta(\lceil n_m/R \rceil)} \Pr\{\boldsymbol{x} \in \mathcal{X}^{\lceil n_m/R \rceil} : \\ &\qquad (\lceil n_m/R \rceil)^{-1} U_{LZ}(\boldsymbol{x}) > R + \epsilon(\lceil n_m/R \rceil)\}, \end{aligned}$$

where the last inequality comes from Proposition. Combining this and (2) yields the theorem □

† $\lceil x \rceil$ denotes the minimum integer greater than or equal to x.

3. Optimum Code for Unifilar Sources

In this section, we restricted our attention to a unifilar source with a finite alphabet $\mathcal{X} = \{1, 2, \cdots, M\}$ and a generic distribution P. For unifilar sources, the state s_i at time instant i obeys the recursion

$$s_i = \psi(x_i, s_{i-1}),$$

where $f : \mathcal{X} \times \mathcal{S} \to \mathcal{S}$ is a known deterministic mapping. Further, we assume that the initial state s_0 is fixed, and clearly we can reconstruct the state sequence s recursively by observing the output sequence $\boldsymbol{x}$. Hence, for a fixed initial state s_0, the probability of a given source sequence $\boldsymbol{x} = x_1 x_2 \cdots x_n$ is given by

$$P(\boldsymbol{x}) = \prod_{i=1}^{n} P(x_i | s_{i-1}).$$

For unifilar sources, Merhav and Neuhoff [3] have shown that the optimum V-F length code provides the exponential decay rate of the probability that the E-CR exceeds R. Combining this result and Theorem 1, we conclude that the LZ code with a fixed codelength can universally achieve the optimum exponential decay rate as the codelength tends to infinity. In what follows, we demonstrate that another simple V-F length code, namely, an adaptive arithmetic code with fixed codelength can also achieve the optimum exponential decay rate.

First, we shall describe the encoding algorithm.

Fundamental encoding algorithm $f_n : \mathcal{X}^* \to \{0,1\}^n$

1) (Initialize) Let $X \leftarrow 0$, $Y \leftarrow 1$, $i \leftarrow 1$, $s \leftarrow s_0$, $u(\tilde{s}) \leftarrow M$ $(\forall \tilde{s} \in \mathcal{S})$ and $F_j(\tilde{s}) = j$ $(j = 0, 1, \cdots, M;\ \forall \tilde{s} \in \mathcal{S})$.

2) (Encoding of the ith symbol) When $x_i = k$, set

$$\begin{aligned} X &\leftarrow X + Y * F_{k-1}(s)/u(s), \\ Y &\leftarrow Y * (F_k(s) - F_{k-1}(s))/u(s), \\ L &\leftarrow \lceil -\log_2 Y + \log_2(i + M) \rceil. \end{aligned}$$

If $L \leqq n$ then go to step 3). Otherwise find an n-bit binary fraction z such that $z \in [X, X + Y)$, and output z, and terminate the algorithm.

3) Set $F_j(s) \leftarrow F_j(s) + 1$ $(j = k, k+1, \cdots, M)$ and $u(s) \leftarrow u(s) + 1$. Then, set $s \leftarrow \psi(x_i, s)$ and $i \leftarrow i + 1$. Go to step 2). □

Next we shall describe the decoding algorithm.

Fundamental decoding algorithm $\varphi_n : \{0,1\}^n \to \mathcal{X}^*$

1) (Initialize) Let $X \leftarrow 0$, $Y \leftarrow 1$, $i \leftarrow 1$, $s \leftarrow s_0$, $u(\tilde{s}) \leftarrow M$ $(\forall \tilde{s} \in \mathcal{S})$ and $F_j(\tilde{s}) = j$ $(j = 0, 1, \cdots, M;\ \forall \tilde{s} \in \mathcal{S})$. Further, let Z be the n-bit fractional number corresponding to the codeword.

2) (Decoding of the ith symbol) Find an index $k \in \{1, 2, \cdots, M\}$ such that $X + Y * F_{k-1}(s)/u(s) \leqq Z < X + Y * F_k(s)/u(s)$, and output k as the ith symbol. Set

$$\begin{aligned} X &\leftarrow X + Y * F_{k-1}(s)/u(s), \\ Y &\leftarrow Y * (F_k(s) - F_{k-1}(s))/u(s), \\ L &\leftarrow \lceil -\log_2 Y + \log_2(i + M) \rceil. \end{aligned}$$

If $L \leqq n$ then go to step 3). Otherwise terminate the algorithm.

3) Set $F_j(s) \leftarrow F_j(s) + 1$ $(j = k, k+1, \cdots, M)$ and $u(s) \leftarrow u(s) + 1$. Then, set $s \leftarrow \psi(x_i, s)$ and $i \leftarrow i + 1$. Go to step 2). □

Remark 1: The above algorithm is an adaptive arithmetic coding using Laplace's law of succession (see e.g. [13], [14]) for the estimation of the generic distribution P of the unifilar source. It should be noted that the variable L is introduced in order to check if the next input symbol can be encoded into the same codeword.

Next theorem shows the asymptotic performance of the proposed codes.

Theorem 2: The probability that the ECR of the proposed code $f_n : \mathcal{X}^* \to \{0,1\}^n$ exceeds R satisfies

$$\begin{aligned} &\liminf_{n\to\infty} \left[-\frac{1}{n} \log_2 \Pr(\rho_{f_n}(X) > R) \right] \\ &\geqq \frac{1}{R} \min_{Q : H(Q) \geq R} D(Q \parallel P), \end{aligned} \tag{3}$$

where the minimum is taken over all probability mass functions Q over $\mathcal{X} \times \mathcal{S}$, and

$$H(Q) \triangleq -\sum_{s\in\mathcal{S}} \sum_{x\in\mathcal{X}} Q(x,s) \log_2 Q(x|s),$$

$$D(Q \parallel P) \triangleq \sum_{s\in\mathcal{S}} \sum_{x\in\mathcal{X}} Q(x,s) \log_2 \frac{Q(x|s)}{P(x|s)},$$

with $Q(x|s) = Q(x,s) / \sum_{s\in\mathcal{S}} Q(x,s)$.

Merhav and Neuhoff [3] have shown that any sequence $\{f_n\}$ of V-F length code f_n with 2^n codewords satisfies

$$\begin{aligned} &\limsup_{n\to\infty} \left[-\frac{1}{n} \log_2 \Pr(\rho_{f_n}(X) > R) \right] \\ &\leqq \frac{1}{R} \min_{Q : H(Q) \geq R} D(Q \parallel P). \end{aligned}$$

Hence, the proposed code can universally achieve the optimum exponential decay rate, as the codelength n tends to infinity. It should be noted that Merhav and Neuhoff have also constructed a code achieving the optimum exponential decay rate [3]. However, their code is based on the enumerative code and is not practical from the viewpoint of complexity for encoding and decoding.

Proof of Theorem 2: For a sequence $\boldsymbol{x} \in \mathcal{X}^m$, $x \in \mathcal{X}$ and $s \in \mathcal{S}$, let

$$q_{\boldsymbol{x}}(x,s) = \frac{1}{m}\sum_{i=1}^{m} \delta(x_i = x, s_{i-1} = s),$$

where $\delta(x_i = x, s_{i-1} = s)$ is the indicator function for $x_i = x$ jointly with $s_{i-1} = s$. Also, let $q_{\boldsymbol{x}}(s) = \sum_{x\in\mathcal{X}} q_{\boldsymbol{x}}(x,s)$ and

$$q_{\boldsymbol{x}}(x|s) = \begin{cases} q_{\boldsymbol{x}}(x,s)/q_{\boldsymbol{x}}(s); & q_{\boldsymbol{x}}(s) > 0 \\ 0 & q_{\boldsymbol{x}}(s) = 0. \end{cases}$$

We denote by $Q_{\boldsymbol{x}}$ the empirical distribution

$$Q_{\boldsymbol{x}} \stackrel{\triangle}{=} \{q_{\boldsymbol{x}}(x,s) : x \in \mathcal{X}, s \in \mathcal{S}\}.$$

Consider a sequence $\boldsymbol{x} \in \mathcal{X}^m$ of empirical distribution $Q_{\boldsymbol{x}}$. If $\boldsymbol{x}$ (and its following symbols) can be encoded into one codeword, the value of Y after reading $\boldsymbol{x}$ satisfies

$$\begin{aligned} Y &= \prod_{s\in\mathcal{S}} \frac{(M-1)!\prod_{j=1}^{M}(mq_{\boldsymbol{x}}(j,s))!}{(mq_{\boldsymbol{x}}(s)+M-1)!} \\ &= \prod_{s\in\mathcal{S}} \binom{mq_{\boldsymbol{x}}(s)+M-1}{M-1}^{-1} \frac{\prod_{j=1}^{M}(mq_{\boldsymbol{x}}(j,s))!}{(mq_{\boldsymbol{x}}(s))!} \\ &\geqq (m+M-1)^{-(M-1)|\mathcal{S}|}\cdot\exp\{-mH(Q_{\boldsymbol{x}})\}. \quad (4) \end{aligned}$$

Then, the probability that a sequence in $\mathcal{X}^m$ is encoded into a sequence of more than mR bits can be upper bounded by

$$\begin{aligned} &\Pr\{\boldsymbol{x} \in \mathcal{X}^m : \ell(f(\boldsymbol{x})) \geqq mR\} \\ &\leqq \sum_{\substack{Q\in\mathcal{P}_m: \\ \lceil -\log_2 Y + \log_2(m+M)\rceil > mR}} \exp\{-mD(Q \,\|\, P)\} \\ &\leqq (m+1)^{M|\mathcal{S}|}\exp\{-\min_{\substack{Q\in\mathcal{P}_m: \\ -\log_2 Y+\log_2(m+M)+1\geqq mR}} mD(Q\,\|\,P)\}, \end{aligned} \quad (5)$$

where $\mathcal{P}_m$ denotes the set of all empirical distributions for $\mathcal{X}^m$, and the last inequality comes from $|\mathcal{P}_m| \leqq (m+1)^{|\mathcal{X}||\mathcal{S}|}$. By using (4), the condition $-\log_2 Y + \log_2(m+M) + 1 \geqq mR$ implies

$$H(Q) + \eta_m \geqq R, \quad (6)$$

where

$$\begin{aligned} \eta_m \stackrel{\triangle}{=} m^{-1}\{&(M-1)|\mathcal{S}|\log_2(m+M-1) \\ &+\log_2(m+M)+1\}. \end{aligned}$$

By choosing $n = mR$ and combining (5) and (6), we obtain

$$\begin{aligned} &-\frac{1}{n}\log_2 \Pr(\rho_{f_n}(X) > R) \\ &= -\frac{1}{n}\log_2 \Pr\{\boldsymbol{x} \in \mathcal{X}^{n/R} : \ell(f(\boldsymbol{x})) > n\} \\ &\geqq -\frac{M|\mathcal{S}|\log(n/R+1)}{n} \\ &\quad +\frac{1}{R}\min_{\substack{Q\in\mathcal{P}_m: \\ H(Q)+\eta_m\geqq R}} D(Q\,\|\,P), \end{aligned}$$

which implies (3). □

Acknowledgment

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) under Grant JSPS-RFTF97P00601.

Appendix: Construction of F-V length code from V-F length code

For the completeness, we describe a method introduced in [3].

1) Let T_n be a complete $|\mathcal{X}|$-ary tree corresponding to a given V-F length code with $N(\leqq 2^n)$ codewords. All words X_i with length $\ell(X_i) \geqq n/R$ can be shortened to $\lceil n/R\rceil$, by pruning all subtrees with roots at depth $\lceil n/R\rceil$. Then, we have a modified tree T_n' with all words no longer than $\lceil n/R\rceil$, and with probability $\Pr\{\ell(X) < n/R\}$ is equal to that of the original code T_n.

2) Every word X with length $\ell(X) < n/R$ is extended to $\lceil n/R\rceil$ by all $|\mathcal{X}|^{\lceil n/R\rceil - \ell(X)}$ possible suffixes, and accordingly, the n-bit codeword for this word is also extended by all possible $((\lceil n/R\rceil - \ell(X))\lceil\log\mathcal{X}\rceil)$-bit suffixes. Then, we have a F-V length code $\tilde{C}_n$ with block length $\lceil n/R\rceil$ and with length function $\tilde{L}(X)$.

Note that the event $\ell(X) < n/R$ for T_n is equivalent to the event $\tilde{L}(X) > n$ for $\tilde{C}_n$.

References

[1] J. Ziv: "Variable-to-fixed length codes are better than fixed-to-variable length codes for Markov sources," IEEE Trans. on Inform. Theory, vol.36, pp.861-863 (1990).

[2] T. J. Tjalkens and F. M. J. Willems: "A universal variable-to-fixed source code based on Lawrence's algorithm," IEEE Trans. on Inform. Theory, vol.38, pp.247-253 (1992).

[3] N. Merhav and D. L. Neuhoff: "Variable-to-fixed length codes provide better large deviations performance than fixed-to-variable length codes," IEEE Trans. on Inform. Theory, vol.38, pp.135-140 (1992).

[4] J. Ziv and A. Lempel: "Compression of individual sequences via variable-rate coding," IEEE Trans. on Inform. Theory, vol.IT-24, pp.530-536 (1978). algorithm for enumerative coding," Trans. IEICE, vol.J80-A, no.3, pp.573-575 (1997).

[5] R. C. Pasco: "Source coding algorithm for fast data compression," Ph.D. Dissertation submitted to Stanford University (1976).

[6] J. Rissanen and G. Langdon: "Arithmetic coding," IBM J.

Res. & Dev., vol.23, pp.149-169 (1979).

[7] M. Guazzo: "A general minimum-redundancy source coding algorithm," IEEE Trans. on Inform. Theory, vol.IT-26, pp.15-25 (1980).

[8] C. B. Jones: "An efficient coding system for long source sequence," IEEE Trans. on Inform. Theory, vol.IT-27, pp.280-291 (1981).

[9] J. Teuhola and T. Raita: "Arithmetic coding into fixed-length codewords," IEEE Trans. on Inform. Theory, vol.40, pp.219-223 (1994).

[10] K. Visweswariah and S. Kulkarni: "Universal variable-to-fixed length source codes for Markov sources," Proc. IEEE ISIT '98, p.48 (1998).

[11] B. Fitingof and Z. Waksman: "Fused trees and some new approaches to source coding," IEEE Trans. on Inform. Theory, vol.34, no.3, pp.417-424 (1988).

[12] N. Merhav: "Universal coding with minimum probability of codeword length overflow," IEEE Trans. on Inform. Theory, vol.37, pp.556-563 (1991).

[13] J. G. Cleary and I. H. Witten: "A comparison of enumerative and adaptive codes," IEEE Trans. on Inform. Theory, vol.IT-30, pp.306-315.

[14] I. H. Witten and T. C. Bell: "The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression," IEEE Trans. on Inform. Theory, vol.37, pp.1085-1094 (1991).

Tomohiko Uyematsu received the B. E., M. E. and D. E. degrees from Tokyo Institute of Technology in 1982, 1984 and 1988, respectively. From 1984 to 1992, He was with the Department of Electrical and Electronic Engineering of Tokyo Institute of Technology, first as a Research Associate, next as a Lecturer, and lastly as an Associate Professor. From 1992 to 1997, he was with School of Information Science of Japan Advanced Institute of Science and Technology as an Associate Professor. Since 1997, he returned to the Department of Electrical and Electronic Engineering of Tokyo Institute of Technology as an Associate Professor. In 1992 and 1996, he was a visiting researcher at the Centre National de la Recherche Scientifique, France and Delft University of Technology, The Netherlands, respectively. He received Shinohara Memorial Young Engineer Award in 1989, and the Best Paper Award in 1993 and 1996, both from IEICE. His current research interests are in the areas of information theory, especially Shannon theory and multi-terminal information theory. Dr. Uyematsu is a member of IEEE.