Joint Source-Channel Coding with Variable Length Codes

Vladimir B. Balakirsky
Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld 1, Germany
on leave from the Data Security Association "Confident", 193060 St.-Petersburg, Russia

We consider an information transmission scheme where a stream of messages is encoded using a variable length code, and the code sequence, obtained as a concatenation of codewords, is transmitted over a noisy discrete memoryless channel. The decoder constructs a stream of estimates of the messages. Note that the problems that concern the use of variable length codes for transmission over noisy channels also appear if the code is used only for source encoding since any channel coding scheme cannot reduce the level of noise up to zero. We propose a convolutional-type decoding algorithm assuming that : (1) the total length of the transmitted code sequence is known to the decoder and (2) decoding delay is allowed and the decoder constructs a maximum likelihood estimate of the whole stream of messages.

One of the basic notions that should be specified, when variable length codes are used for noisy channels, is the event called "decoding error". We define this event as follows. Suppose, $m_1, m_2, ...$ is a stream of messages to be transmitted and let $l(m_1), l(m_2), ...$ be a sequence of lengths of the corresponding codewords. Then the transmitted data can be represented as a sequence of pairs $(m_1, t_1), (m_2, t_2), ...$, where $t_k = t_{k-1} + l(m_{k-1}), k = 1, 2, ...$, and $t_0 = 0$. The decoder also represents the decoded stream $\hat{m}_1, \hat{m}_2, ...$ as a sequence of pairs $(\hat{m}_1, \hat{t}_1), (\hat{m}_2, \hat{t}_2), ...$, where $\hat{t}_k = \hat{t}_{k-1} + l(\hat{m}_{k-1}), k = 1, 2, ...$, and $\hat{t}_0 = 0$. We say that the k-th decoding error takes place if (m_k, t_k) is not included into the decoded sequence of pairs. Thus, the decoder is interested both in the messages and time instants when they were transmitted.

For example, consider the information transmission scheme which we call "decoding of noisy answers when the questions are unknown". Suppose that there are boxes of M different types, all boxes of the m-th type have lengths $l(m)$ meters. Somebody forms a sequence of boxes putting them on a moving line whose speed is 1 meter per second. Every second an asking machine sends a question to the line, i.e., presents a subset of the set $\{1, ..., M\}$, and the line answers 1 if the type of a current box belongs to this subset and 0 otherwise. This answer goes to the decoder and can be corrupted by noise. We also assume that the asking machine gets a special signal at the moment when a new box appears. Then we can construct a binary variable length code consisting of M codewords $\mathbf{x}_1, ..., \mathbf{x}_M$ of lengths $l(1), ..., l(M)$ and form the sets S_l consisting of the messages m such that $l(m) \geq l$ and the l-th component of $\mathbf{x}_l$ is equal to 1; $l = 1, ..., l_{max} = \max\{l(1), ..., l(M)\}$. At every second $t = 1, 2, ...$ the asking machine presents a subset $S_{f(t)}$, where $f(t) = 1$ if the symbol for changing of a box is arrived and $f(t) = f(t-1) + 1$ otherwise ($f(1) = 1$). Then the decoder receives a sequence of codewords of the variable length code and can recognize both the types of the boxes and the time instants when these boxes appeared with small error probability if the parameters of the code match channel characteristics.

The maximum likelihood decoder can be viewed as a version of the Viterbi decoder for convolutional codes [1]. It contains l_{max} buffers and stores an estimate of the code sequence of length $t - l$ and its metric (the logarithm of the conditional probability to receive a given sequence if the stored code sequence was transmitted) in the l-th buffer before the t-th step; at the t-th step the decoder forms an estimate of the code sequence of length t and its metric, $l = 1, ..., l_{max}$. The contents of the buffers can be recurrently determined for $t = 1, ..., T$, where T is a given length of the code sequence, and the decoder outputs the sequence of messages corresponding to the code sequence formed at time T. If the codewords of the variable length code consist of the symbols assigned to the edges of a non-uniform tree (note that any code can be represented in this way), then the decoder looks for a maximum likelihood path in the trellis containing N nodes at every level, where N is the total number of non-terminal nodes of the tree. Note also that the assumption concerning a known T is not important and, including a truncation step into the decoding procedure [1], we can decode each transmitted code symbol with a fixed delay.

Let $\lambda_t(m)$ denote the probability of the k-th decoding error provided that the sequence of messages is fixed in such a way that the message $m_k = m$ is transmitted at time t. A key observation of the analysis of $\lambda_t(m)$ is the fact that the total number of code sequences having length l and obtained as a concatenation of codewords of the variable length code can be upper-bounded by 2^{rl}, where the parameter r satisfies the inequality $\sum_m 2^{-rl(m)} \leq 1$. As a result, we prove that the expectation of $\lambda_t(m)$ taken over the ensemble of random variable length codes with a given length distribution can be upper-bounded by

$$\overline{\lambda}_t(m) \leq K(l_{min}, r) 2^{-l(m) E_{ran}(r)},$$

where $E_{ran}(r)$ is the value of the random coding bound for the channel and $K(l_{min}, r)$ is a non-exponential function of $l_{min} = \min\{l(1), ..., l(M)\}$ and r.

The bound above has some interesting properties, and we present them for the Bernoulli source : $M = 2^n$ and $P_m = p^{w(m)}(1-p)^{n-w(m)}$ is the probability that the message $m \in \{1, ..., M\}$ will be generated; $w(m)$ denotes the Hamming weight of a binary representation of the integer $m - 1$. Assign a variable length code with $l(m) = n$ if $w(m) \leq n(p - \varepsilon_n)$ and $l(m) = -\log P_m / h(p)$ otherwise, where $h(p)$ is the binary entropy function. Then ε_n can be specified in such a way that the ratio of the average code length to n tends to 1 as $n \to \infty$, and an exponential term of the upper bound on the average decoding error probability, $\frac{1}{T} \sum_{t=1}^{T} \sum_m P_m \cdot \overline{\lambda}_t(m)$, is equal to $2^{-nE_{ran}(h(p))}$. If the random coding bound gives an exact exponent for the channel coding at rate $h(p)$, then this bound is asymptotically tight.

References

[1] A.J.Viterbi and J.K.Omura, *Principles of Digital Communication and Coding,* New York : McGraw-Hill, 1979.