論文 84-21-6-15

An Upper Bound on Run-Length Coding Entropy

(Run-Length 符號化의 Entropy 上限에 관한 研究)

宋 翊 鎬*, 安 秀 桔*

(Iick Ho Song, and Souguil ANN)

要 約

二進 run-length 符號化에서 run-length의 最大값이 M으로 制限되었을 때의 最大 entropy를 구하는 過程을 보였고, 이때 M이 無限大가 되면 Huang의 上限[1]이 된다.

Abstract

A method of calculating the maximum entropy per run in binary run-length coding is given when run-length is limited to a maximum of M. And Huang's bound[1] can be obtained from the present result as M tends to infinity.

I. Introduction

In [1], an upper bound on the entropy per run in binary run-length coding was given as

$$H_{max} = R \ln R - (R-1) \ln(R-1) \tag{1}$$

where R is the average run-length and it was assumed that quantized run-length could take any integer value from unity to infinity. In this correspondence, the maximum value of the quantized run-length r is limited to M. We use "ln" for natural logarithm.

II. The Maximum Entropy

As usual,

$$p_i = \Pr(r=i) \quad i=1, 2, \ldots, M \tag{2}$$

*正會員, 서울대학교 공과대학 전자공학과

(Dept. Electron. Eng., Seoul National Univ.)

接受日字：1984年 5月 19日

$$\sum_{i=1}^{M} p_i = 1 \tag{3}$$

$$1 < R = \bar{r} = \sum_{i=1}^{M} ip_i < M \tag{4}$$

$$H(p) = -\sum_{i=1}^{M} p_i \ln p_i \tag{5}$$

It is assumed that $M > 1$. To use the Lagrange multiplier method we define

$$F(p) = H(p) + a\left(\sum_{i=1}^{M} ip_i - R\right) + b\left(\sum_{i=1}^{M} p_i - 1\right) \tag{6}$$

One easily obtains that

$$P_i = \exp(ai+b-1) \tag{7}$$

for maximum entropy and so

$$H(p)\big|_{max} = 1 - b - aR . \tag{8}$$

From (3) and (7)

$$1 = (1-x^M)e^{a+b-1}/(1-x)$$

or

$$b = 1 - \ln x + \ln [(1-x)/(1-x^M)] \qquad (9)$$

where

$$x = e^a, \ a \neq 0 \qquad (10)$$

From (4), (7) and (10),

$$R = [(1-x^M) - Mx^M(1-x)] / [(1-x)(1-x^M)]$$

or

$$(M-R)x^{M+1} - (M-R+1)x^M + Rx + 1 - R = 0. \qquad (11)$$

Theorem 1. Equation (11) has one and only one solution in the interval $0 < x < 1$ when $R < (M+1)/2$ and, in the interval $x > 1$ when $R > (M+1)/2$.

Proof:

Let

$$f(x) = (M-R)x^{M+1} - (M-R+1)x^M + Rx + 1 - R \qquad (12)$$

then

$$f(o) = 1 - R < 0 \qquad (13)$$
$$f(1) = 0 \qquad (14)$$
$$f'(o) = R > 1 \qquad (15)$$
$$f'(1) = 0 \qquad (16)$$

and

$$f''(x) = M(M+1)(M-R)x^{M-2}(x-x_1) \qquad (17)$$

where

$$0 < x_1 = 1 + (2R-M-1)/((M+1)(M-R)). \qquad (18)$$

In the case $R < (M+1)/2$

$$0 < x_1 < 1 \qquad (19)$$

and so $f'(x)$ has unique minimum at $x = x_1$ in the interval $0 < x < 1$, which with (15) and (16), means there is a unique solution to $f'(x)=0$ in that interval. This fact with (13) and (14), in turn, states that $f(x)=0$ has a unique solution in the same interval.

In the case $R > (M+1)/2$

$$x_1 > 1 \qquad (20)$$

and as in the previous case, we see that $f(x)=o$ has a unique solution in the interval $x > 1$. Finally, in the case $R = (M+1)/2$, we can obtain

$$P_i = 1/M \qquad i = 1, 2, \ldots, M \qquad (21)$$

$$H(p)|_{max} = \ln M = \ln(2R-1) \qquad (22)$$

from (3) and (7).

Next we investigate the behavior of maximum entropy as M tends to infinity.

In the case $R < (M+1)/2$, the first two terms

Table 1. Maximum entropy for M and R. ($M > 2R-1$)

R	M	Hmax
2.0	4	1.2839
	6	1.3675
	10	1.3853
	20	1.3863
	35	1.3863
	∞	1.3863
5.0	6	1.3675
	10	2.2874
	30	2.5007
	50	2.5020
	80	2.5020
	∞	2.5020
10.0	15	2.5985
	30	3.1904
	50	3.2452
	100	3.2508
	170	3.2508
	∞	3.2508
50.0	60	3.3475
	100	4.6050
	500	4.9019
	700	4.9020
	900	4.9020
	∞	4.9020
100.0	150	4.8450
	500	5.5929
	1000	5.6001
	1800	5.6002
	2000	5.6002
	∞	5.6002
500.0	600	5.6074
	1500	7.1399
	4000	7.2133
	6000	7.2136
	9000	7.2136
	∞	7.2136

of (11) are negligible because $0 < x < 1$. Thus the solution is

$$x = 1-1/R$$

and

$$H(p)|_{max} = R \ln R - (R-1) \ln (R-1).$$

In the case $R > (M+1)/2$, the last three terms of (11) are negligible because $x > 1$. Thus the solution is $x=1$, which gives exactly the same results as in (21) and (22). Thus

$$H(p)|_{max} \qquad \infty$$

In the case $R = (M+1)/2$ one easily sees that

$$H(p)|_{max} \qquad \infty$$

from (22).

III. Discussion

Theorem I eanables us to obtain the maximum entropy from (8), (9), (10) and (11), i.e., if M and R is given we can solve (11) for x by numerical method, calculate a and b from (10) and (9) respectively and obtain the maximum entropy from (8). Some of the results are shown in Table 1.

From the results above, we can see that Huang's bound is rapidly approached when M is sufficiently large.

Acknowledgement

We express our appreciation to the referees and Miss Taeyoung Seong for their constructive suggestions and to Mr. In Kyung Kim and Mr. Moon Ho Song for their sincere help.

References

[1] T.S. Huang, "An upper bound on the entropy of run-length coding", *IEEE. Tr. Infor. Theory*, vol. IT-20, pp. 675-676, Sep., 1974.

[2] 송익호, 안수길, "RLC (Run-Length Coding) 의 entropy 상한에 관한 연구", 대한전자공학회, 추계종합학술대회, vol.2, no.2, 1983.