Efficient Run-Length Encoding of Binary Sources with Unknown Statistics

Max H. M. Costa
Universidade Estadual de Campinas
FEEC/DECOM, Caixa Postal 6101
13081-970 Campinas, SP, Brazil

Henrique S. Malvar
Microsoft Research
One Microsoft Way
Redmond, WA 98052, USA

Binary entropy coders are used in many multimedia codec standards; an example is JPEG-LS, which uses adaptive Golomb-Rice coders. In this work we consider a binary memoryless source whose most probable symbol is zero, occurring with probability $\theta \geq ½$, and θ is unknown. Typical Golomb coders restrict the Golomb parameter $M = 2^k$ (a.k.a. Rice coding). In this work we propose the uses of two subsets of Golomb parameters: $M = 2^k$ ($h = 0$) or $M = 3 \cdot 2^{k-1}$ ($h = 1$), as well as an extra $k = 0$, $h = 1$ mode, as an extension to [1]. We derive simple polynomials whose roots can be used to determine precisely the optimal values of the parameters k and h. as a function of θ. For example, the crossover probabilities for which we should switch from using $M' = 2^k$ to using $M'' = 3 \cdot 2^{(k-1)}$ are given by $\theta'_k = x_1^{2/M'}$, $k = 1,2,3,\ldots$, where $0 < x_1 < 1$ is the only sensible solution to $x^3 + x^2 = 1$, that is $x_1 = 0.754877666247$.

Using a maximum-likelihood approach, we derive a nearly-optimal adaptation strategy, which uses a simple backward adaptation rule for k and h. While Golomb-Rice coders with $M = 2^k$ have an excess rate with respect to the source entropy of up to 4.2% for binary sources with unknown statistics, the proposed encoder has an excess rate of less than 2%, as shown in Figure 1. The full paper is available in [2].

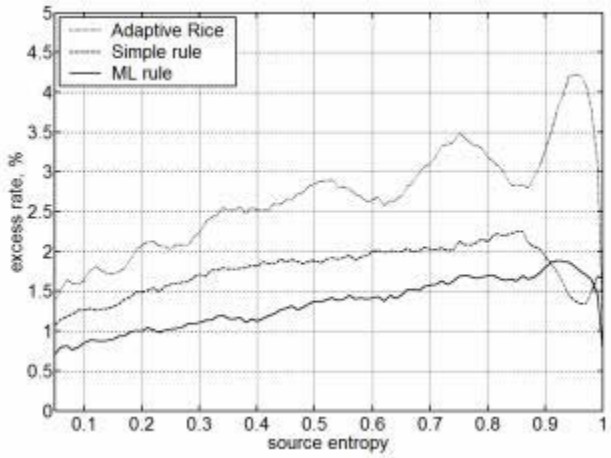

Figure 1. Excess rate for the proposed Golomb-like coder, compared to an adaptive Rice coder, for stationary sources with unknown θ.

References

[1] H. Tanaka and A. Leon-Garcia, "Efficient run-length encodings," *IEEE Trans. Information Theory*, vol. IT-28, pp. 880–890, Nov. 1982

[2] M. H. M. Costa and H. S. Malvar, "Efficient run-length encoding of binary sources with unknown statistics," Microsoft Research Tech. Report TR-2003-95, Dec. 2003.

Proceedings of the Data Compression Conference (DCC'04)
1068-0314/04 $ 20.00 © 2004 IEEE