On the Inadequacy of Golomb-Rice Codes for Adaptive Coding

Amir Said (said@ieee.org)
Hewlett Packard Labs, Palo Alto, CA, USA

Complex data sources, like images and audio, require sophisticated coding contexts and source modeling. Fortunately, in many cases the high cost for estimating a very large number of conditional probabilities and then computing optimal codes, can be avoided by storing sets of codewords, and selecting the best choice based on local source estimates. Golomb-Rice prefix codes are commonly used for such purpose because of their convenient features. They are parameterized with a single integer m (or k if $m = 2^k$), are easy to implement, and are optimal for sources with geometric distribution ($p_s = (1-\rho)\rho^s$), which is a good approximation of those found in practice.

In this work [1], we consider the fact that, even when the source distribution *is* geometric, the Golomb-Rice codes are truly optimal only when the source parameter ρ is known with certainty, which in practice is never the case. We investigate how these codes perform—on sources with geometric distribution—depending on how ρ is estimated from previous samples.

Next, we analyze the possible changes in the code to increase their robustness, but keeping the useful structural properties. We maintain the unary+binary structure, but consider sets of codewords parameterized by infinite sequences $(m_0, m_1, m_2, \ldots)$. While this is similar to other proposals for generalization of Golomb-Rice codes, we use this form for theoretical analysis, i.e., the intention is not to propose a "new" type of code for particular applications, but to observe how the optimal codes change with different models of source uncertainty and estimation methods.

The new codes can be designed by updating the maximum-likelihood or Bayesian estimation of ρ, during the unary part, or by designing optimal codes using marginal probabilities. We show that data for modeling the source uncertainty can be accurately computed with numerical integration, or using incomplete Beta functions, and we develop an algorithm, capable of dealing with the infinite number of symbols, to quickly find the optimal codes with the desired structure. It is based on implicit enumeration, using some convenient upper and lower bounds on expected rate, which guarantee quick convergence to optimal codes within a chosen precision.

Numerical results show that the optimal codes are, as expected, always better than Golomb-Rice codes. The codes designed solely using estimation of ρ, on the other hand, are not even always better than Golomb-Rice codes. This shows the importance of integrating the source estimation parameters into the code design, and the identification of mappings from source-sample data directly to code parameters.

[1] A. Said, *Determination of optimal parameterized prefix codes for adaptive coding,* HP Labs Report, Palo Alto, CA, March 2005 (http://www.hpl.hp.com/techreports/).

Proceedings of the 2005 Data Compression Conference (DCC'05)
1068-0314/05 $ 20.00 **IEEE**