New results on Length Restricted Prefix Codes

Ruy Luiz Milidiu, Eduardo Sany Laber, Artur Alves Pessoa.
Deparamento de Informática, PUC-RJ, Brasil
{milidiu,laber,artur}@inf.puc-rio.br

Given an alphabet $\Sigma = \{a_1, \ldots, a_n\}$ with an associated list of weights $W = [w_1, \ldots, w_n]$, an optimal prefix code for Σ is a prefix code that minimizes $\sum_{i=1}^{n} w_i l_i$, where l_i is the length of the code assigned to symbol a_i. An optimal length-restricted prefix codes also minimizes $\sum_{i=1}^{n} w_i l_i$ with the additional constraint that $l_i < L$ for $i = 1, \ldots, n$, where $L \geq \lceil \log n \rceil$.

Recently, we proposed the WARM-UP algorithm that constructs suboptimal length-restricted codes [1]. This algorithm is based on a Lagrangean relaxation of the length restricted prefix code problem. WARM-UP looks for the minimum value of x such that the Huffman tree constructed for the modified set of weights $W = [w_1^x, \ldots, w_n^x]$, where $w_i^x = max\{x, w_i\}$, has its height not greater than the length restriction L. This algorithm has $O(nL \log\log(n/L))$ and $O(n \log w_n)$ implementations, spending $O(n)$ space in both cases. Practical experiments with the WARM-UP show that it often obtains an optimal length-restricted prefix code in linear time, given that he list of weights is already sorted.

One motivation for implementing length-restricted codes is to improve the decoding speed for very large alphabets. Nevertheless, length-restricted codes introduce a compression degradation. It has been observed that this degradation is very small in practical cases. We show [2] that the difference between the average code length of an optimal length restricted code and that of an unrestricted optimal prefix code is bounded above by $1/\psi^{L-\lceil \log n \rceil -1}$, where ψ is the golden ratio 1.618. The proof of this bound is constructive and leads us to a new algorithm for constructing length restricted prefix codes. This algorithm runs in $O(n)$ time if the input list of weights is already sorted. Furthermore, this algorithm is $1/\psi^{L-\lceil \log n \rceil -1}$-approximative.

[1] **Milidiú, R.L. and Laber, E.S.,** *The WARM-UP ALGORITHM: A Lagrangean Construction of Length Restricted Huffman Codes*, **Monografias em Ciência da Computação 15/96, Departamento de Informática, PUC-RJ, 1996. Submitted to SIAM Journal on Computing.**

[2] **Milidiú, R.L., and Laber, E.S.,** *Bounding the Inefficiency Introduced by Length-Restricted Prefix Codes*, **Monografias em Ciência da Computação 33/97, Departamento de Informática, PUC-RJ, 1997. Submitted to Journal of the ACM.**