On the Cost of Worst-Case Coding Length Constraints

Dror Baron and Andrew C. Singer

Abstract

We investigate the redundancy that arises from adding a worst-case length-constraint to uniquely decodable fixed to variable codes over achievable Huffman codes. This is in contrast to the traditional metric of the redundancy over the entropy. We show that the cost for adding constraints on the worst-case coding length is small, and that the resulting bound is related to the Fibonacci numbers.

Keywords

Data compression, Fibonacci numbers, Huffman coding, redundancy, source coding, uniquely decodable.

I. Introduction

A fundamental tradeoff in lossless source coding is that some inputs can be compressed only if others are expanded. A reasonable objective is to compress well on average, while expanding little in the worst-case.

The tradeoff between the expected coding length and the worst-case coding expansion has received research attention. In [1] an algorithm for finding a code meeting these constraints is proposed, and in [2] the redundancy of the expected coding length over the entropy is bounded. In this paper, we investigate the redundancy of the expected coding length of constrained codes over that of achievable Huffman codes [3]. We bound this redundancy by a term that decays exponentially in the worst-case coding expansion, and note that this term is related to the Fibonacci numbers. The problem is stated in Section II, the main results are given in Section III, and a discussion is provided in Section IV.

II. Problem formulation

Consider a discrete alphabet $\mathcal{X}$ and length-N input sequences x, i.e., $x \in \mathcal{X}^N$. We define a *source code* C as a mapping $C : \mathcal{X}^N \to \mathcal{X}^*$ where $\mathcal{X}^*$ is the set of finite-length sequences over $\mathcal{X}$. Following [4], let $C(x)$ be the *codeword* corresponding to x, and $l(x)$

This work was supported in part by NSF grants No. MIP-97-07633 and NSF CDA 96-24396.

The authors are with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign. They can be reached at dbaron@uiuc.edu and acsinger@uiuc.edu.

the *length* of $C(x)$. The *expected coding length* $L(C)$ for a random variable X with a probability mass function (PMF) $p(x)$ is defined as

$$L(C) \triangleq \sum_{x \in \mathcal{X}^N} p(x) l(x). \tag{1}$$

A *uniquely decodable* code is a source code C with a dual mapping $\tilde{C} : \mathcal{X}^* \to \mathcal{X}^N$, such that $\tilde{C}(C(x)) = x, \forall x \in \mathcal{X}^N$. We say that the code C *expands* x when $l(x) > N$, and we define the *worst-case coding expansion* as

$$W(C) \triangleq \max_{x \in \mathcal{X}^N} \{l(x) - N\} \tag{2}$$

(it is non-negative because no source code compresses all input sequences). Given the PMF $p(x), x \in \mathcal{X}^N$, and an (integer) constraint Δ on the worst-case coding expansion, the *constrained expected coding length* is defined as

$$L_\Delta^* \triangleq \min_{C:\ W(C) \leq \Delta} \{L(C)\}. \tag{3}$$

When the constraint is relaxed, and Δ is increased, the constrained expected coding length will go down until at some stage it is equal to the unconstrained expected coding length. This is the expected coding length of the Huffman code, and is denoted by L_∞^*. In the following section we bound $L_\Delta^* - L_\infty^*$.

III. Results

Any Huffman code can be viewed as a tree, where codewords correspond to *leaves*, and their prefixes correspond to *internal nodes*. The *nodes*, be they leaves or internal nodes, make up a *Huffman tree*. A leaf corresponds to some input x and its probability $p(x)$; an internal node corresponds to a set of descendant leaves, and the total probability of those leaves. Lemma 1 bounds the probabilities corresponding to internal nodes, and is related to Theorem 7 in [5], which bounds the probabilities corresponding to leaves. Both proofs are by induction, the only difference is due to initial conditions.

Lemma 1: Any depth-k internal node in a Huffman tree corresponds to a set of codewords with total probability p_k satisfying

$$p_k \leq \frac{1}{f_k}, \tag{4}$$

where $f_{n+1} \triangleq f_n + (|\mathcal{X}| - 1) f_{n-1}$, with initial conditions $f_0 \triangleq 1,\ f_1 \triangleq 2 - \frac{1}{|\mathcal{X}|}$.

Proof: In the Huffman tree, let $p_k, p_{k-1}, \ldots, p_0 = 1$ be the probabilities corresponding to internal nodes on the path from our depth-k node, denoted by α, to the root. Let $q_l^i,\ i \in \{1, \ldots, |\mathcal{X}| - 1\}$ be the corresponding probability of node i merged with p_l into p_{l-1}. We prove $p_l \geq f_{k-l} p_k$ by induction on l. First, $p_k = 1 \cdot p_k = f_0 p_k$. Second, the lemma requires α to be an internal node, so at least one of its descendant nodes corresponds to at least a probability of $\frac{1}{|\mathcal{X}|}p_k$. But α and its parent are internal nodes, so when the parent is created in the Huffman algorithm, its descendants are nodes corresponding to the minimal probabilities among all the nodes at that stage, so $q_k^i \geq \frac{1}{|\mathcal{X}|}p_k$. Therefore,

$$\begin{aligned} p_{k-1} &= p_k + \sum_{i=1}^{|\mathcal{X}|-1} q_k^i && (5) \\ &\geq p_k + (|\mathcal{X}|-1)\left(\frac{p_k}{|\mathcal{X}|}\right) \\ &= f_1 p_k, && (6) \end{aligned}$$

where (5) is the merging of corresponding probabilities. The inductive step is

$$\begin{aligned} p_{l-1} &= p_l + \sum_{i=1}^{|\mathcal{X}|-1} q_l^i && (7) \\ &\geq p_l + (|\mathcal{X}|-1)p_{l+1} && (8) \\ &\geq \left(f_{k-l} + (|\mathcal{X}|-1)\, f_{k-l-1}\right) p_k && (9) \\ &= f_{k-l+1} p_k && (10) \end{aligned}$$

where (7) is similar to (5), the reasoning in the second step leads to (8), (9) is by induction, and (10) uses the definition of f. The lemma (4) follows because $p_0 = 1$. ■

For $|\mathcal{X}| = 2$, f_n are essentially scaled Fibonacci numbers, i.e., $f_0 = \frac{2}{2}, f_1 = \frac{3}{2}, f_2 = \frac{5}{2}, f_3 = \frac{8}{2}, \cdots$. In the general case, the recursion for f_n can be solved using methods for difference equations [6] leading to

$$\begin{aligned} f_n &= \left(\frac{1}{2} + \frac{\frac{3}{2} - \frac{1}{|\mathcal{X}|}}{\sqrt{4|\mathcal{X}|-3}}\right)\left(\frac{1+\sqrt{4|\mathcal{X}|-3}}{2}\right)^n \\ &+ \left(\frac{1}{2} - \frac{\frac{3}{2} - \frac{1}{|\mathcal{X}|}}{\sqrt{4|\mathcal{X}|-3}}\right)\left(\frac{1-\sqrt{4|\mathcal{X}|-3}}{2}\right)^n, \ n \geq 0. && (11) \end{aligned}$$

Theorem 1: Given a random variable X with a PMF $p(x)$ over $\mathcal{X}^N$, then for $\Delta > 0$

$$L_\Delta^* - L_\infty^* \leq \frac{1}{f_{\Delta-1}}. \tag{12}$$

Proof: We begin with C_H, a Huffman code that achieves L_∞^*. We create a new code $\tilde{C}$ (not necessarily a Huffman code), with length function $\tilde{l}(x)$ and expected coding length $L(\tilde{C})$, by modifying C_H in two steps. Step 1: prune $X_1 = \{x : \ l(x) \geq N + \Delta\}$ from the Huffman tree of C_H. If X_1 is empty we are done, else nodes were pruned off a depth-$N+\Delta-1$ internal node, so there exists a depth-$\Delta-1$ internal node, including the one in the path from the depth-$N+\Delta-1$ internal node to the root. Step 2: take any depth-$\Delta-1$ internal node in the tree; denote it by β. Let X_2 be the descendant leaves of β. Replace β with a new node γ with two descendants. The first descendant is β, and the second descendant is a depth-N full tree with up to $|\mathcal{X}|^N$ leaves that can accommodate all of X_1, since $|X_1| < |\mathcal{X}|^N$. The full tree for X_1 starts at depth Δ and goes up to depth $N+\Delta$, so $\tilde{l}(x) \leq N+\Delta,\ \forall x \in X_1$. For X_2, $l(x) < N+\Delta,\ \forall x \in X_2$, so adding an additional symbol gives $\tilde{l}(x) \leq N+\Delta,\ \forall x \in X_2$. Therefore $W(\tilde{C}) \leq \Delta$. The structure β originally resided at depth $\Delta-1$ in the tree, so by Lemma 1 $\sum_{x\in X_2} p(x) \leq \frac{1}{f_{\Delta-1}}$. Therefore, $\tilde{C}$ satisfies

$$\begin{aligned}
L(\tilde{C}) &= \sum_{x\in\mathcal{X}^N} p(x)\tilde{l}(x) \\
&\leq \sum_{x\in X_1} p(x)\tilde{l}(x) + \sum_{x\in X_2} p(x)\tilde{l}(x) + \sum_{x\notin X_1\cup X_2} p(x)\tilde{l}(x) \\
&\leq \sum_{x\in X_1} p(x)l(x) + \sum_{x\in X_2} p(x)(l(x)+1) + \sum_{x\notin X_1\cup X_2} p(x)l(x) \qquad (13)\\
&= \sum_{x} p(x)l(x) + \sum_{x\in X_2} p(x) \\
&\leq L(C_H) + \frac{1}{f_{\Delta-1}}, \qquad (14)
\end{aligned}$$

where (13) arises because codewords in X_1 became shorter. The result is obtained by noting that $L(C_H) = L^*_\infty$. □

IV. Discussion

We begin with several technical remarks on Theorem 1. First, the theorem does not apply for $\Delta = 0$, because there is no depth-$\Delta-1$ internal node that can be split. In fact, for $\Delta = 0$ there is no expansion, nor is there any compression, thus $L^*_0 = N$. Second, the theorem upper bounds $L^*_\Delta - L^*_\infty$, but we cannot give a lower bound, because for a uniform PMF we have $L^*_\Delta = N$, hence $L^*_\Delta - L^*_\infty = 0$. Third, we can get a stronger bound on the expected coding length for $\Delta = 2$. In this case, there always exists some depth-$\Delta-1$ node (not necessarily internal) with $\sum_{x\in X_2} p(x) \leq \frac{1}{|\mathcal{X}|}$, so

$$L^*_2 - L^*_\infty \leq \frac{1}{|\mathcal{X}|}. \qquad (15)$$

Although Lemma 1 bounds the probabilities corresponding to depth-k internal nodes, there could be nodes at that depth that correspond to even smaller probabilities.

The constructive method used in the proof of the theorem can be used to derive codes that satisfy constraints on the worst-case coding expansion, but these are not necessarily optimal codes. However, the theorem is useful because it bounds the cost of the constraint by a term that decays exponentially in the expansion. A tighter bound in the main theorem could be obtained by finding a stronger version of Lemma 1 for the depth-k node that corresponds to the smallest probabilities.

Acknowledgments

The authors wish to thank the two anonymous reviewers and Marcelo Weinberger for their insightful comments.

References

[1] A. Moffat, A. Turpin, and J. Katajainen, "Space-efficient construction of optimal prefix codes", *Proc. Data Compression Conference*, Snowbird, UT, pp. 192-201, March 1995.

[2] R. M. Capocelli and A. De Santis, "On the Redundancy of Optimal Codes with Limited Word Length", *IEEE Trans. Information Theory*, vol. IT-38, no. 2, pp. 439-445, March 1992.

[3] D. A. Huffman, "A Method for the Construction of Minimum Redundancy Codes", *Proc. IRE*, vol. 40, no. 9, pp. 1098-1101, September 1952.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York; John Wiley and Sons, 1991.

[5] R. M. Capocelli and A. De Santis, "A Note on D-ary Huffman Codes", *IEEE Trans. Information Theory*, vol. IT-37, no. 1, pp. 174-179, January 1991.

[6] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing.* Englewood Cliffs, NJ; Prentice Hall, 1989.