# Abstract Huffman Coding and PIFO Tree Embeddings

**Keri D'Angelo★ Dexter Kozen➔**
**Cornell University**
**Computer Science Department**
**Ithaca, New York 14853-7501, USA**
★kd349@cornell.edu
➔kozen@cs.cornell.edu

K. D'Angelo and D. Kozen, "Abstract Huffman coding and PIFO tree embeddings," Tech. Rep. https://arxiv.org/abs/2301.02878, Cornell University, January 2023.

# Introduction

- A similarity exists between Huffman coding and a recent algorithm developed for compiling PIFO (priority-in first-out) trees to trees of fixed shape, but they work with different underlying algebraic operations.

- We utilize the monad of $d$-ary prefix codes on the category Set; that is, $\mathcal{C} : \mathsf{Set} \to \mathsf{Set}$ where $\mathcal{C}X$ is the set of pairs $(C, r)$ such that $C$ is a prefix code over a $d$-ary alphabet for a fixed $d \geq 2$ and $r : C \to X$ and $\mathcal{C}h : \mathcal{C}X \to \mathcal{C}Y$ is defined by $\mathcal{C}h(C, r) = (C, h \circ r)$ for $h : X \to Y$.

# Applications

- For Huffman coding, we want to minimize the value $\sum_{x \in C} |x| \cdot r(x)$, where $r(x)$ is the frequency of letters assigned to the codeword $x$.
- For PIFO tree embeddings, we wish to minimize the value $max_{x \in C} |x| + r(x)$, where $r(x)$ is the height of the subtree.

# Algorithm

Suppose we are given a multiset $M$ of weights in $W$, $|M| \geq 2$. We would like to find an optimal tree for this multiset of weights. The following is a recursive algorithm to find such an optimal tree.

1. Say there are $n \geq 2$ elements in $M$. Let $k \in \{2, \ldots, d\}$ such that $n \equiv k \bmod (d-1)$. Let $a_0, \ldots, a_{k-1}$ be the $k$ elements of least weight. Form the object

$$(\{0, 1, \ldots, k-1\}, i \mapsto a_i) \in \mathcal{C}W.$$

   If there are no other elements of $M$, return that object.

2. Otherwise, let

$$M' = \{(\{0, 1, \ldots, k-1\}, i \mapsto a_i)\} \cup \{\eta_W(a) \mid a \in M \setminus \{a_0, \ldots, a_{k-1}\}\},$$

   a multiset of $n - k + 1 < n$ elements of $\mathcal{C}W$.

3. Recursively call the algorithm at step 1 with $M'' = \{\mathcal{C}w(E, t) \mid (E, t) \in M'\}$, a multiset of elements of $W$. This returns a tree $(D, s)$ of type $\mathcal{C}W$ that is optimal for $M''$. The bijective map $s : D \to M''$ factors as $\mathcal{C}w \circ s'$ for some bijective $s' : D \to M'$, and $(D, s') \in \mathcal{C}^2 W$.

   Flatten this to $\mu_W(D, s') \in \mathcal{C}W$ and return that value.

# Application 1- Huffman Coding

- The multiset of weights comes from $W = \mathbb{R}_+ = \{a \in \mathbb{R} \mid a \geq 0\}$ with weighting $w(C, r) = \sum_{x \in C} r(x)$.

- $(W, w)$ is an Eilenberg Moore algebra for the $d - ary$ prefix code monad $(\mathcal{C}, \mu, \eta)$.

- The order $\leq$ on $\mathcal{C}W$ can be defined as follows- Let $\alpha : \mathcal{C}W \to W$ be defined as $\alpha(C, r) = \sum_{x \in C} |x| \cdot r(x)$ and define $(C, r) \leq (D, s)$ if $(C, r) \sim (D, s)$; that is, there is a bijective map $f : C \to D$ such that $r = s \circ f$ and $\alpha(C, r) \leq \alpha(D, s)$.

**Theorem 1.** *The algorithm shown previously for the algebra $(\mathbb{R}_+, w)$ and ordering relation $\leq$ is equivalent to Huffman's algorithm and produces an optimal Huffman code for a given multiset of weights.*

# Application 2- PIFO Tree Embeddings

- The multiset of weights comes from $W = \mathbb{N}$ with weighting $w(C, r) = max_{x \in C}|x| + r(x)$.
- $(W, w)$ is an Eilenberg Moore algebra for the $d - ary$ prefix code monad $(\mathcal{C}, \mu, \eta)$.
- We define $(C, r) \leq (D, s)$ if there is a bijective function $f : C \to D$ such that $r = s \circ f$ and $w(C, r) \leq w(D, s)$.

**Theorem 2.** *The algorithm shown previously for the algebra $(\mathbb{N}, w)$ and ordering relation $\leq$ is equivalent to determining whether an embedding of a PIFO tree into a bounded d-ary tree exists and finding the embedding if so.*

# References


[1] A. Sivaraman, S. Subramanian, M. Alizadeh, S. Chole, S.-T. Chuang, A. Agrawal, H. Balakrishnan, T. Edsall, S. Katti, and N. McKeown, "Programmable packet scheduling at line rate," in *SIGCOMM*, 2016.

[2] A. Alcoz, A. Dietmüller, and L. Vanbever, "SP-PIFO: Approximating push-in first-out behaviors using strict-priority queues," in *NSDI*, 2020.

[3] A. Mohan, Y. Liu, N. Foster, T. Kappé, and D. Kozen, "Formal abstractions for packet scheduling," Tech. Rep. `http://arxiv.org/abs/2211.11659`, Cornell University, November 2022.

[4] K. D'Angelo and D. Kozen, "Abstract Huffman coding and PIFO tree embeddings," Tech. Rep. `https://arxiv.org/abs/2301.02878`, Cornell University, January 2023.