UNIVERSITY OF LIVERPOOL

Department of Computer Science
Sebastian Wild

Date: 2020-03-25
Version: 2020-03-11 17:30

Tutorial 7 for COMP 526 – Applied Algorithmics, Winter 2020

Problem 1 (Move-to-front transform)

Let $S = (20, 30, 30, 20, 40, 30, 20, 20, 20)$ be an input sequence of numbers whose values are initially stored in the list $Q = [20, 30, 40]$. Build an output sequence and trace the content of Q throughout the execution of *MTF (Move-to-Front) algorithm*.

Problem 2 (Lempel-Ziv-Welch compression)

Given word $w =$ `ASNXASNASNA` over the ASCII character set (relevant parts of ASCII are provided on the right). Construct, step by step, the Lempel-Ziv-Welch (LZW) factorization of w (i.e., the phrases encoded by one codeword) and provide the compressed representation of w; it suffices to show the encoded text C using integer numbers (no need for binary encodings).

Code	Character
65	`A`
...	...
78	`N`
...	...
83	`S`
...	...
88	`X`
...	...

Problem 3 (No Free Lunch)

Prove the following *no-free-lunch* theorems for lossless compression.

1. *Weak version:* For every compression algorithm A and $n \in \mathbb{N}$ there is an input $w \in \Sigma^n$ for which $|A(w)| \geq |w|$, i. e. the "compression" result is no shorter than the input.

   Hint: Try a proof by contradiction. There are different ways to prove this.

2. *Strong version:* For every compression algorithm A and $n \in \mathbb{N}$ it holds that

   $$\left|\{w \in \Sigma^{\leq n} : |A(w)| < |w|\}\right| < \tfrac{1}{2} \cdot \left|\Sigma^{\leq n}\right| .$$

   In words, less than half of all inputs of length at most n can be compressed below their original size.

   Hint: Start by determining $|\Sigma^{\leq n}|$.

The theorems hold for every non-unary alphabet, but you can restrict yourself to the binary case, i.e., $\Sigma = \{\texttt{0}, \texttt{1}\}$.

We denote by $\Sigma^\star$ the set of all (finite) strings over alphabet Σ and by $\Sigma^{\leq n}$ the set of all strings with size $\leq n$. As domain of (all) compression algorithms, we consider the set of (all) *injective* functions in $\Sigma^\star \to \Sigma^\star$, i.e., functions that map any input string to some output string (encoding), where no two strings map to the same output.