2018 Fall Data Compression Homework #1

EE 248583
Advisor: 電子所 高立人 副教授

106368002 張昌祺 Justin, Chang-Qi Zhang
justin840727@gmail.com
Due Date: November 12 2018

Problem 1 Entropy

Let X be a random variable with an alphabet $H = \{1, 2, 3, 4, 5\}$. Please determine $H(X)$ for the following three cases of probability mass function $p(i) = prob[X = i]$. (15%)

(a) $P(1) = P(2) = 1/2$:

Ans

Firstly we take a look the result images which are generated by my program for both Lena and baboon gray-level resolution from 8 bits to 1 bit.

(b) $P(i) = 1/4, for\ i = 1, 2, 3, and\ p(4) = p(5) = 1/8$:

Ans

(c) $P(i) = 2^{-i}, for\ i = 1, 2, 3, 4, and\ p(5) = 1/16$:

Ans

Problem 2 Huffman Code

Design a Huffman code C for the source in Problem 1. (15%)

(a) Specify your codewords for individual pmf model in Problem 1.

Ans

(b) Compute the expected codeword length and compare with the entropy for your codes in (a).

Ans

(c) Design a code with minimum codeword length variance for the pmf model in Problem 1.(b)

Ans

Problem 3 Empirical Distribution C++

Empirical distribution. In the case a probability model is not known, it can be estimated from empirical data. Let's say the alphabet is $H = \{1, 2, 3, ... , m\}$. Given a set of observations of length N , the empirical distribution is given by $p = total\ number\ of\ symbol\ 1/N,\ for\ i = 1, 2, 3, ..., m$. Please determine the empirical distribution for **santaclaus.txt**, which is an ASCII file with only lower-cased English letters (i.e., $a \sim z$), space and CR (carriage return), totally 28 symbols. The file can be found on the class web site. Compute the entropy. (14%)

Ans

Problem 4 Huffman Code Encode C++

Write a program that designs a Huffman code for the given distribution in Problem 3. (14

Ans

Problem 5 Adaptive Huffman Tree

Let X be a random variable with an alphabet H , i.e., the 26 lower-case letters. Use adaptive Huffman tree to find the binary code for the sequence **a a b b a**. (24%)
You are asked to use the following 5bits fixed-length binary code as the initial codewords for the 26 letters. That is
a: 00000
b: 00001
⋮
z: 11001
Note: Show the Huffman tree during your coding process.

Ans

Problem 6 Golomb Encoding and Decoding.

(a) Find the Golomb code of n=21 when m=4

(b) Find the Golomb code of n=14 when m=4

(c) Find the Golomb code of n=21 when m=5

(d) Find the Golomb code of n=14 when m=5

(e) A two-integer sequence is encoded by Golomb code with m=4 to get the bitstream 11101111000. What's the decoded two-integer sequence?

(f) A two-integer sequence is encoded by Golomb code with m=5 to get the bitstream 11101111000 (the same bitstream as that in (e)). What's the decoded two-integer sequence?
Hint: The unary code for a positive integer q is simply q 1s followed by a 0.