Problems 5 - Dictionary coders, Test transforms, Quantization

Coding Technology

Illés Horváth

2025/11/14

Dictionary codes, Text transforms

Reminder: Dictionary coders.

LZ77: lookup buffer, search buffer, search for longest match, add one character, record: (position, length of match, new character).

LZ78: add one character to a string already seen before, record: (address of old string, novelty).

Reminder: Text transforms.

MTF transforms: move current character to the front of the alphabet after each step.

Burrows–Wheeler transform. Transform: order all cyclic permutations alphabetically, codeword is last column + the position of the original message in the alphabetical ordering. Inverse transform: starting from last and first column, match sections of increasing length in each step.

Problem 1

Compute the next two records of the LZ77 algorithm, starting from the following position.

| a b l a k b a r n a | a b l a k b a a h a r | a r r a d a h r. . .

Solution.

- first step:

  10

  | a b l a k b a r n a | a b l a k b a a h a r | a r r a d a h r. . . (10,7,a)

  7

- next step:

  a b l a k b a r | n a a b l a k b a a | h a r a r r a d a h r | . . . (0,0,h)

Problem 2

An LZ77 coder has parameters $h_s = h_\ell = 5$. Decode the following sequence of records: (0,0,C), (0,0,A), (2,1,D), (3,2,B), (2,5,A).

Solution. (0,0,X)-type records code a single character (X), so the first two records are decoded to CA.

The numbers in the record (2,1,D) are decoded as a copy command: move back 2 positions and copy 1 character (C). Then the D is added to the end $\rightarrow$ CACD.

Similarly, (3,2,B) results in CACDACB.

For the record (2,5,A), we execute a similar copy command, but we only have 2 characters (CB) to copy; in this case, we keep copying those 2 characters cyclically until we get 5 characters (CBCBC), then add an A to the end $\rightarrow$ CACDACBCBCBCA.

Problem 3

Using LZ78, encode the message AABABBBABBABB. Also convert the dictionary to binary.

Solution. We use A $\rightarrow$ 0, B $\rightarrow$ 1 fixed length character encoding.

A|A B|A B B|B|A B B A|B B

1	(0,A)	0
2	(1,B)	011
3	(2,B)	011101
4	(0,B)	011101001
5	(3,A)	0111010010110
6	(4,B)	01110100101101001

Problem 4

Apply MTF transform to the text “tobeornottobe”.

Solution.

abcdefghijklmnopqrst	tobeornottobe	20
tabcdefghijklmnopqrs	tobeornottobe	20,16
otabcdefghijklmnpqrs	tobeornottobe	20,16,4
botacdefghijklmnpqrs	tobeornottobe	20,16,4,7
ebotacdfghijklmnpqrs	tobeornottobe	20,16,4,7,3
oebtacdfghijklmnpqrs	tobeornottobe	20,16,4,7,3,19
roebtacdfghijklmnpqs	tobeornottobe	20,16,4,7,3,19,16
nroebtacdfghijklmpqs	tobeornottobe	20,16,4,7,3,19,16,3
onrebtacdfghijklmpqs	tobeornottobe	20,16,4,7,3,19,16,3,6
tonrebacdfghijklmpqs	tobeornottobe	20,16,4,7,3,19,16,3,6,1
tonrebacdfghijklmpqs	tobeornottobe	20,16,4,7,3,19,16,3,6,1,2
otnrebacdfghijklmpqs	tobeornottobe	20,16,4,7,3,19,16,3,6,1,2,6
btonreacdfghijklmpqs	tobeornottobe	20,16,4,7,3,19,16,3,6,1,2,6,6

Problem 5

Compute the Burrows–Wheeler transform of the text "andanand".
Solution.

andanand		anandand	
ndananda		andanand	$\leftarrow$
danandan		andandan	
anandand	$\rightarrow$	danandan	
nandanda		dandanan	
andandan		nandanda	
ndandana		ndananda	
dandanan		ndandana	

The Burrows–Wheeler transform is ("ddnnnaaa", 2).

Problem 6

Apply inverse Burrows–Wheeler transform to (“NNBAAA”, 4).

Solution.

ABANAN
ANABAN
ANANAB
BANANA $\leftarrow$
NABANA
NANABA

The inverse transform is “BANANA”.

Quantization

Reminder. Quantization: round values from a larger set to values from a smaller set.

A quantizer $Q : \mathbb{R} \to \mathbb{R}$ is described by levels $x_1, \ldots, x_N$ and domains $B_1, \ldots, B_N$.

Real random signal X. Distortion: $D(Q) = \mathbb{E}((X - Q(X))^2)$.

Nearest neighbour rule, center of gravity rule, Lloyd-Max property.

Optimal quantizer: explicit derivation only when N is very small.

Otherwise, Lloyd-Max algorithm for iterative approximation:

1. Initialize levels.
2. Update domains using nearest neighbour rule.
3. Compute $D(Q)$, if the decrease is below threshold, stop.
4. Update values using center of gravity rule, repeat.

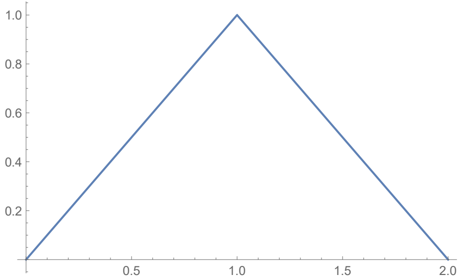
Problem 7

The probability density function of the random signal X is

$$f(x) = \begin{cases} 1 - |x-1| & \text{for } x \in [0,2] \\ 0 & \text{otherwise} \end{cases}$$

Compute the optimal 3-level quantizer for X explicitly. (Hint: use symmetry.)

Solution. The pdf looks like this:



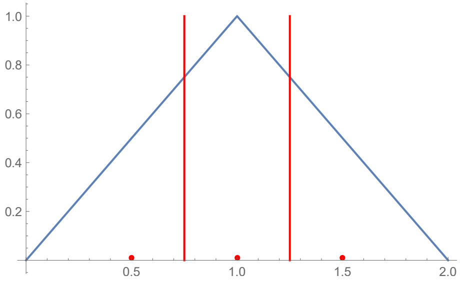
Problem 7

Let the boundaries be $y \in [0,1]$ and $2-y \in [1,2]$, so

$$B_1 = [0, y] \qquad B_2 = [y, 2-y] \qquad B_3 = [2-y, 2]$$

According to the center of gravity rule (and the symmetry), the levels are

$$x_1 = \frac{\int_0^y x \cdot x \mathrm{d}x}{\int_0^y x \mathrm{d}x} = \frac{2y}{3}, \qquad x_2 = 1, \qquad x_3 = 2 - \frac{2y}{3}.$$



Problem 7

$$D(Q) = \int_0^y (x - 2y/3)^2 x \mathrm{d}x + \int_y^1 (x-1)^2 x \mathrm{d}x +$$
$$\int_1^{2-y} (x-1)^2 (2-x) \mathrm{d}x + \int_{2-y}^2 (x - (2 - 2y/3))^2 (2-x) \mathrm{d}x =$$
$$-\frac{4}{9}y^4 + \frac{4}{3}y^3 - y^2 + \frac{1}{6}$$

$$\frac{\mathrm{d}}{\mathrm{d}y}\left(-\frac{4}{9}y^4 + \frac{4}{3}y^3 - y^2 + \frac{1}{6}\right) = -\frac{16}{9}y^3 + 4y^2 - 2y = 0$$
$$y\left(-\frac{16}{9}y^2 + 4y - 2\right) = 0$$
$$y_1 = 0 \quad y_2 = 0.75 \quad y_3 = 1.5$$

Problem 7

Is $y_2 = 0.75$ a minimum?

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}\left(-\frac{16}{3}y^4 + \frac{4}{3}y^3 - y^2 + \frac{1}{6}\right)\bigg|_{y=y_2} > 0$$

so y_2 is a minimum and the optimal quantizer is:

$$B_1 = [0, 0.75] \quad B_2 = [0.75, 1.25] \quad B_3 = [1.25, 2]$$

$$x_1 = 0.5 \qquad x_2 = 1 \qquad x_3 = 1.5$$

Problem 7

Solution 2. Once we have $x_1 = \frac{2y}{3}$, y can also be obtained from the nearest neighbour rule: $y = \frac{x_1+x_2}{2}$ gives a linear equation for y that can be solved:

$$y = \frac{x_1 + x_2}{2} = \frac{\frac{2y}{3} + 1}{2} \quad \rightarrow \quad y = 0.75,$$

and the rest follows.

Problem 8

We want to find the optimal 3-level quantization for $X \sim U([0,2])$. Execute two iterations of the Lloyd-Max algorithm, starting from the levels $x_1 = 0, x_2 = 0.4, x_3 = 1.6$.

Solution. Step 1: initialization. $x_1 = 0, x_2 = 0.4, x_3 = 1.6$.

Step 2: compute the domains.

$$B_1 = [0, 0.2] \qquad B_2 = [0.2, 1.0] \qquad B_3 = [1.0, 2.0]$$

Step 3:

$$D(Q) = \int_0^{0.2} (x - 0.2)^2 \cdot \frac{1}{2} \mathrm{d}x + \int_{0.2}^{1.0} (x - 0.4)^2 \cdot \frac{1}{2} \mathrm{d}x + \\ + \int_{1.0}^{2} (x - 1.6)^2 \cdot \frac{1}{2} \mathrm{d}x \approx 0.0805$$

Step 4: update the levels.

$$x_1 = 0.1 \qquad x_2 = 0.6 \qquad x_3 = 1.5$$

Problem 8

Second iteration.

Step 2: update the domains.

$$B_1 = [0, 0.35] \qquad B_2 = [0.35, 1.05] \qquad B_3 = [1.05, 2.0]$$

Step 3:

$$D(Q) = \int_0^{0.35} (x - 0.1)^2 \cdot \frac{1}{2} \mathrm{d}x + \int_{0.35}^{1.05} (x - 0.6)^2 \cdot \frac{1}{2} \mathrm{d}x + \\ + \int_{1.05}^{2} (x - 1.5)^2 \cdot \frac{1}{2} \mathrm{d}x \approx 0.0665$$

Step 4: update the levels.

$$x_1 = 0.175 \qquad x_2 = 0.7 \qquad x_3 = 1.525$$

Problem 9

What is the optimal 3-level quantization for $X \sim U([0, 2])$?
Compute the distortion.

Solution. For uniform distributions, the optimal quantizer is always discrete uniform, that is, all domains have the same length:

$$B_1 = [0, 2/3] \qquad B_2 = [2/3, 4/3] \qquad B_3 = [4/3, 2]$$

$$D(Q) = \int_0^{2/3} (x - 1/3)^2 \cdot \frac{1}{2} \mathrm{d}x + \int_{2/3}^{4/3} (x - 1)^2 \cdot \frac{1}{2} \mathrm{d}x + \\ + \int_{4/3}^{2} (x - 5/3)^2 \cdot \frac{1}{2} \mathrm{d}x = 2/27 \approx 0.0370$$