Coding and Information Theory
Chapter 2
Optimal Codes

Xuejun Liang

2019 Fall

Chapter 2: Optimal Codes

- Need to balance between using words which are
  - long enough to allow effective decoding, and
  - short enough for economy

- Optimal codes
  - the instantaneous codes with least average word length

Content of Chapter 2

2.1 Optimality

2.2 Binary Huffman Codes

2.3 Average Word-length of Huffman Codes

2.4 Optimality of Binary Huffman Codes

2.5 r-ary Huffman Codes

2.6 Extensions of Sources

2.1 Optimality

- Let S be a source and assume that the probabilities

$$p_i = \Pr(X_n = s_i) = \Pr(s_i)$$

where

$$0 \leq p_i \leq 1, \qquad \sum_{i=1}^{q} p_i = 1.$$

- Assume code C for S has word-lengths $l_1, l_2, \ldots l_q$. Then the Average Word-Length is defined as

$$L = L(C) = \sum_{i=1}^{q} p_i l_i \ .$$

- Given r and the probability distribution (p_i), we try to find instantaneous r-ary codes C minimizing $L(C)$.
  - Such codes are called optimal or compact codes

Optimality (Cont.)

- Example 2.1
  - Let S be the daily weather (as in Example 1.2)
  - with $p_i = \frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ for $i = 1, 2, 3$.
  - Consider two instantaneous codes
  - binary code $\mathcal{C} : s_1 \mapsto 00,\ s_2 \mapsto 01,\ s_3 \mapsto 1$
  - $L(C) = ?$

  - binary code $\mathcal{D} : s_1 \mapsto 00,\ s_2 \mapsto 1,\ s_3 \mapsto 01$
  - $L(D) = ?$

Optimality (Cont.)

- Lemma 2.2
  - Given a source S and an integer r, the set of all average word-lengths $L(C)$ of uniquely decodable r-ary codes C for S is equal to the set of all average word-lengths $L(C)$ of instantaneous r-ary codes C for S.
- Definition
  - An instantaneous r-ary code C is defined to be optimal if $L(C) = L_{min}(S)$, which is the greatest lower bound of average word-lengths.
- Theorem 2.3
  - Each source S has an optimal r-ary code for each integer $r \geq 2$.

2.2 Binary Huffman Codes

- Let $T = Z_2 = \{0,1\}$, Given a source S, we renumber the source-symbols $s_1, \dots, s_q$, so that

$$p_1 \geq p_2 \geq \cdots \geq p_q \,.$$

- Form a reduced source S' by combining the two least-likely symbols.
- Given any binary code C' for S', we can form a binary code C for S:

$$\begin{array}{lllll} \mathcal{S}: & s_1, & \dots, & s_{q-2}, & \underbrace{s_{q-1},\ s_q} \\ \mathcal{S}': & s_1, & \dots, & s_{q-2}, & s' \end{array}$$

$$\begin{array}{llll} p_1, & \dots, & p_{q-2}, & \underbrace{p_{q-1},\ p_q} \\ p_1, & \dots, & p_{q-2}, & p' \end{array}$$

$$\begin{array}{lllll} \mathcal{C}: & w_1, & \dots, & w_{q-2}, & \underbrace{w'0,\ w'1} \\ \mathcal{C}': & w_1, & \dots, & w_{q-2}, & w' \end{array}$$

Binary Huffman Codes (Cont.)

- Lemma 2.4
  - If the code C' instantaneous then so is C.
- Huffman code for S
  - Constructed by

$$\mathcal{S} \to \mathcal{S}' \to \cdots \to \mathcal{S}^{(q-2)} \to \mathcal{S}^{(q-1)}$$

$$\mathcal{C} \leftarrow \mathcal{C}' \leftarrow \cdots \leftarrow \mathcal{C}^{(q-2)} \leftarrow \mathcal{C}^{(q-1)}.$$

  - Note: $C^{(q-1)} = \{\varepsilon\}$ and $C^{(q-2)} = \{\varepsilon 0, \varepsilon 1\} = \{0, 1\}$
  - It is instantaneous
- Example 2.5
  - Let S have q = 5 symbols s_1,…,s_5 with probabilities $p_i =$ 0.3, 0.2, 0.2, 0.2, 0.1. Compute Huffman code and $L(C)$

How the probability distribution affects the average word-length of Huffman codes

- Example 2.6
  - Let S have q = 5 symbols s_1,…,s_5 again, but now suppose that they are equiprobable, that is, $p_1 = \ldots = p_5 = 0.2$. Compute Huffman code and $L(C)$.
- In general, the greater the variation among the probabilities p_i, the lower the average word-length of an optimal code.
- Note: **entropy** can be used to measure the amount of variation in a probability distribution.
  - Will study later in next chapter.

2.3 Average Word-length of Huffman Codes

$$\begin{aligned} L(\mathcal{C}) - L(\mathcal{C}') &= p_{q-1}(l+1) + p_q(l+1) - (p_{q-1} + p_q)l \\ &= p_{q-1} + p_q \\ &= p', \end{aligned} \tag{2.3}$$

- Note p' is the "new" probability created by reducing S to S'.
- If we iterate this, using the fact that $L\left(C^{(q-1)}\right) = |\varepsilon| = 0$, we find that

$$L(C) = p' + p'' + \cdots + p^{(q-1)} \qquad (2.4)$$

- the sum of all the new probabilities $p', p'', \ldots, p^{(q-1)}$ created in reducing S to $S^{(q-1)}$.
  - Try Example 2.5 and Example 2.6

2.4 Optimality of Binary Huffman Codes

- Definition
  - Two binary words w_1 and w_2 to be siblings if they have the form $x0$, $x1$ (or vice versa) for some word $x \in T^*$.
- Lemma 2.7
  - Every source S has an optimal binary code D in which two of the longest code-words are siblings.
- Theorem 2.8
  - If C is a binary Huffman code for a source S, then C is an optimal code for S.

2.5 r-ary Huffman Codes

- If we use an alphabet T with $|T| = r > 2$, then the construction of r-ary Huffman codes is similar to that in the binary case.
  - Merge r source symbols together at a time
  - Note: may need to add some dummy symbols such that
$$q \equiv 1 \bmod (r - 1)$$
- Example 2.9
  - Let q = 6 and r = 3. Since r - 1 = 2 we need q $\equiv$ 1mod (2), so we adjoin an extra symbol s_7 to S, with p_7 = 0
  - The reduction process now gives ……

r-ary Huffman Codes (Cont.)

- Example 2.10
  - Let q = 6 and r = 3 and suppose that the symbols $s_1, \ldots, s_6$, of S have probabilities p_i = 0.3, 0.2, 0.2, 0.1, 0.1, 0.1.
  - After adjoining s_7 with p_7 = 0, we find that the reduction process is as follows:

2.6 Extensions of Sources

- Let S be a source with
  - q symbols $s_1, \dots, s_q$ of
  - probabilities $p_1, \dots, p_q$
- The n-th extension S^n of S is the source with
  - q^n symbols $s_{i_i} \dots, s_{i_n}$ ($s_{i_j} \in S$)
  - probabilities $p_{i_i} \dots, p_{i_n}$
- Note: The probabilities $p_{i_i} \dots, p_{i_n}$ form a probability distribution by
  - Expanding the left-hand side of the equation

$$(p_1 + \cdots + p_q)^n = 1^n = 1$$

Extensions of Sources: Examples

- Example 2.11
  - Let S have source S = $\{s_1, s_2\}$ with p_1 = $2/3$, p_2 =$1/3$.
  - Then S^2 has source alphabet = $\{s_1s_1, s_1s_2, s_2s_1, s_2s_2\}$ with probabilities $4/9$, $2/9$, $2/9$, $1/9$.
- Example 2.12: S is as in Example 2.11
  - A binary Huffman code C: $s_1 \mapsto 0,\ s_2 \mapsto 1$
  - Average word-length $L(C) = 1$
  - Construct a Huffman code C^2 for S^2
  - Average word-length $L(C^2) = ?$
  - You will see $L(C^2)/2 < L(C) = 1$

Extensions of Sources: decoding

- Decode a pair (two consecutive symbols), rather than one symbol, at a time.
  - Not quite instantaneous
  - A bounded delay while waiting for pairs to be completed
- Can construct a Huffman code C^3 for S^3
  - Can show $L(C^3)/3 < L(C^2)/2$
- Continuing this principle, construct a Huffman code C^n for S^n
  - the average word-length $L(C^n)/n \to ?$ as $n \to \infty$