Coding and Information Theory
Chapter 2
Optimal Codes

Xuejun Liang

This is the second lecture of chapter 2

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

2.2 Binary Huffman Codes

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

$$\begin{array}{ll}
\mathcal{S}: & s_1, \ \ldots \ , s_{q-2}, \ \underbrace{s_{q-1}, \ s_q} \\
\mathcal{S}': & s_1, \ \ldots \ , s_{q-2}, \qquad s' \\
 & \\
 & p_1, \ \ldots \ , p_{q-2}, \ \underbrace{p_{q-1}, \ p_q} \\
 & p_1, \ \ldots \ , p_{q-2}, \qquad p' \\
 & \\
\mathcal{C}: & w_1, \ \ldots \ , w_{q-2}, \ \underbrace{w'0, \ w'1} \\
\mathcal{C}': & w_1, \ \ldots \ , w_{q-2}, \qquad w'
\end{array}$$

Example 2.5

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

- Let S have q = 5 symbols $s_1,\ldots,s_5$ with probabilities

  $p_i =$ 0.3, 0.2, 0.2, 0.2, 0.1.

  Compute Huffman code and $L(C)$

S				0.3	0.2	0.2	0.2	0.1
C				01	10	11	000	001
S'			0.3	0.3	0.2	0.2		
C'			00	01	10	11		
S''		.4	0.3	0.3				
C''		1	00	01				
S'''	0.6	.4						
C'''	0	1						
S'''' 1.0								
C'''' ε								

Example 2.6

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

- Let S have $q = 5$ symbols $s_1,\ldots,s_5$ again, but now suppose that they are equiprobable, that is,

  $$p_1 = \ldots = p_5 = 0.2.$$

  Compute Huffman code and $L(C)$.

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

- 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

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

- 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

	S					0.3	0.2	0.2	0.2	0.1
	S'				0.3	0.3	0.2	0.2		
2.5	S''			0.4	0.3	0.3				
	S'''		0.6	0.4						
	S''''	1.0								

	S					0.2	0.2	0.2	0.2	0.2
	S'				0.4	0.2	0.2	0.2		
2.6	S''			0.4	0.4	0.2				
	S'''		0.6	0.4						
	S''''	1.0								

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.
- **Proof**: By Theorem 2.3, there is an optimal binary code for S

Let us choose such a code D which has the minimal total word length σ(D)

Choose a longest code-word w in D

Assume w = x0, then x1 ∈ D. So D has two longest sibling code-words

If x1 ∉ D. Let D' = (D – {x0}) ∪ {x}.

Then D' is a prefix code and σ(D') < σ(D). This is a contradiction!

Theorem 2.8: If C is a binary Huffman code for a source S, then C is an optimal code for S.

- **Proof**: Lemma 2.4 shows that C is instantaneous, so it is sufficient to show that L(C) is minimal

We use induction on the number q of source-symbols.

If q = I then C = {ε} with L(C) = 0, so the result is trivially true.

Assume that L(C) is minimal for all sources with q - 1 symbols

Prove that L(C) is minimal for all sources with q symbols

Let S = { $s_1, s_2, \dots, s_{q-2}, s_{q-1}, s_q$ } and S' = { $s_1, s_2, \dots, s_{q-2}, s'$ }, $s' = s_{q-1} V s_q$

Now let D: $s_i \rightarrow x_i$ be the optimal binary code for S given by Lemma 2.7

D has a sibling pair of longest code-words: $x_{q-1} = x0 \; and \; x_q = x1$

Now form a code D' for S': $s_i \rightarrow x_i$ (i<q-1) and s'→ x

$L(\mathcal{D}) - L(\mathcal{D}') = p_{q-1} + p_q = L(\mathcal{C}) - L(\mathcal{C}')$, $L(\mathcal{D}') - L(\mathcal{C}') = L(\mathcal{D}) - L(\mathcal{C})$

Now C' is a Huffman code for S', a source with q-1 symbols, so by the induction hypothesis C' is optimal

$L(\mathcal{C}) \leq L(\mathcal{D})$

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 ……

Example 2.10

Let q = 6 and r = 3 and suppose that the symbols $s_1, \dots, s_6$, 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:

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:

p_i = 0.3, 0.2, 0.2, 0.1, 0.1, 0.1, 0.0

C = {1,   00, 01, 02, 20, 21, 22}

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 $s_{i_1} \dots, s_{i_n}$ ($s_{i_j} \in S$)
  - probabilities $p_{i_1} \dots, p_{i_n}$
- Note: The probabilities $p_{i_i} \dots, p_{i_n}$ form a probability distribution by
  - Expanding the left-hand side of the equation

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

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 = $\{s_1 s_1, s_1 s_2, s_2 s_1, s_2 s_2\}$
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 $n \to \infty$