Coding and Information Theory Overview
Chapter 1: Source Coding

Xuejun Liang

The first Lecture of Chapter 1

Overview

- Information Theory and Coding Theory are two related aspects of the problem of how to transmit information efficiently and accurately from a source, through a channel, to a receiver.
- Based on Mathematics areas:
  - Probability Theory and Algebra
  - Combinatorics and Algebraic Geometry

Important Problems

- How to compress information, in order to transmit it rapidly or store it economically
- How to detect and correct errors in information

Information Theory vs. Coding Theory

- Information Theory uses probability distributions to quantify information (through the entropy function) , and to relate it to the average word-lengths of encodings of that information
  - In particular, Shannon's Fundamental Theorem Guarantees the existence of good error-correcting codes (ECCs)
- Coding Theory is to use mathematical techniques to construct ECCs, and to provide effective algorithms with which to use ECCs.

Chapter 1: Source Coding

1.1 Definitions and Examples

1.2 Uniquely Decodable Codes

1.3 Instantaneous Codes

1.4 Constructing Instantaneous Codes

1.5 Kraft's Inequality

1.6 McMillan's Inequality

1.7 Comments on Kraft's and McMillan's Inequalities

1.1 Definitions and Examples

- A sequence $s = X_1X_2X_3 \ldots$ of symbols X_n, emitting comes from a source S
- The source alphabet of $S = \{s_1, s_2, \ldots, s_q\}$
- Consider X_n as random variables and assume that
  - they are independent and
  - have the same probability distribution p_i.

$$\Pr(X_n = s_i) = p_i \quad \text{for } i = 1, \ldots, q.$$

$$p_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{q} p_i = 1$$

Examples

- Example 1.1
  - S is an unbiased die, S = {1, . . . , 6} with q = 6, X_n is the outcome of the n-th throw, and p_i = $1/6$.
- Example 1.2
  - S is the weather at a particular place, with X_n representing the weather on day n, S = {good, moderate, bad}.

$$p_1 = 1/4,\ p_2 = 1/2,\ p_3 = 1/4.$$

- Example 1.3
  - S is a book, S consists of all the symbols used, X_n is the n-th symbol in the book, and p_i is the frequency of the i-th symbol in the source alphabet.

Code alphabet, symbol, word

- Code alphabet $T = \{t_1, \ldots, t_r\}$ consisting of r code-symbols t_j.
  - Depends on the technology of the channel
  - Call r the radix (meaning "root“ or “base”)
  - Refer to the code as an r-ary code
  - When r = 2, binary code, T = Z_2 = {0, 1}
  - When r = 3, ternary code, T = Z_3 = {0, 1, 2}
- Code word: a sequence of symbols from T

Encode and Example

- To encode $s = X_1X_2X_3$ ..., we represent $X_n = s_i$ by
  - $s_i \rightarrow w_i$ (its code word)
  - $s \rightarrow t$ (one by one)
  - we do not separate the code-words in t
- Example 1.4
  - If S is an unbiased die, as in Example 1.1, take $T = Z_2$ and let w_i be the binary representation of the source-symbol s_i
    - $s_i = i$ ($i = 1, . . . ,6$)
    - $w_1 = 1, w_2 = 10, w_3 = 11, w_4 = 100, w_5 = 101, w_6 = 110$
  - s = 53214 $\rightarrow$ t =10111101100
  - Could write t = 101.11.10.1.100 for clearer exposition

Define codes more precisely

- A word w in T is a finite sequence of symbols from T, its length $|w|$ is the number of symbols.
- The set of all words in T is denoted by $T*$, including empty word ε.
- The set of all non-empty words in T is denoted by T^+

$$T^* = \bigcup_{n \geq 0} T^n \qquad \text{and} \qquad T^+ = \bigcup_{n > 0} T^n,$$

where $T^n = T \times \cdots \times T$

Define codes more precisely (Cont.)

- A source code (simply a code) C is a function $S \rightarrow T^+$

$$w_i = C(s_i) \in T^+, \qquad i = 1, 2, \ldots, q$$

- Regard C as a finite set of words $w_1, w_2, \ldots, w_q$ in T^+.
- C can be extended to a function $S^* \rightarrow T^*$

$$\mathbf{s} = s_{i_1} s_{i_2} \ldots s_{i_n} \mapsto \mathbf{t} = w_{i_1} w_{i_2} \ldots w_{i_n} \in T^*$$

- The image of this function is the set

$$C^* = \{w_{i_1} w_{i_2} \ldots w_{i_n} \in T^* \mid \text{each } w_{i_j} \in C, \ n \geq 0\}$$

- The average word-length of C is

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

  - $where\ l_i = |w_i|$

Example 1.5

- Recall Example 1.4
  - Source symbols:
    - $s_1 = 1, s_2 = 2, s_3 = 3, s_4 = 4, s_5 = 5, s_6 = 6$
  - Probability distribution
    - $p_1 = \frac{1}{6}, p_2 = \frac{1}{6}, p_3 = \frac{1}{6}, p_4 = \frac{1}{6}, p_5 = \frac{1}{6}, p_6 = \frac{1}{6}$
  - Code words:
    - $w_1 = 1, w_2 = 10, w_3 = 11, w_4 = 100, w_5 = 101, w_6 = 110$
  - Word lengths
    - $l_1 = 1$, $l_2 = l_3 = 2$ and $l_4 = l_5 = l_6 = 3$
- So, average word length

$$L(\mathcal{C}) = \frac{1}{6}(1 + 2 + 2 + 3 + 3 + 3) = \frac{7}{3}.$$

The aim is to construct codes C

a) there is easy and unambiguous decoding t -> s ,
b) the average word-length $L(C)$ is small.

- The rest of this chapter considers criterion (a) , and the next chapter considers (b).

1.2 Uniquely Decodable Codes

- A code C is uniquely decodable (u.d. for short) if each $t \in T^*$ corresponds under C to at most one $s \in S^*$;
  - in other words, the function $C : S^* \rightarrow T^*$ is one-to-one,
- Will always assume that the code-words w_i in C are distinct.
  - Under this assumption, the definition of unique decodability of C is that whenever

$$u_1 \dots u_m = v_1 \dots v_n$$

with $u_1, \dots, u_m, v_1, \dots, v_n \in C$, we have $m = n$ and $u_i = v_i$ for each i.

Example 1.6

- In Example 1.4, the binary coding of a die is not uniquely decodable.
- Give an example.
- Can you fix it?

Theorem 1.7

- If the code-words w_{i} in C all have the same length, then C is uniquely decodable.
  - If all the code-words in C have the same length l, we call C **a block code of length $\boldsymbol{l}$**.

Example: Uniquely Decodable But Not Block Code

- Example 1.8
  - The binary code C given by

$$s_1 \mapsto w_1 = 0,\ s_2 \mapsto w_2 = 01,\ s_3 \mapsto w_3 = 011$$

  - has variable lengths, but is still uniquely decodable.
  - for example,

$$\mathbf{t} = 001011010011 = 0.01.011.01.0.011$$

$$\Rightarrow \quad \mathbf{s} = s_1 s_2 s_3 s_2 s_1 s_3.$$

Definition of C_n and C_∞

- We define
  - $\mathcal{C}_0 = \mathcal{C}$, and
  - $\mathcal{C}_n = \{ w \in T^+ \mid uw = v \text{ where } u \in \mathcal{C},\ v \in \mathcal{C}_{n-1} \text{ or } u \in \mathcal{C}_{n-1},\ v \in \mathcal{C} \}$ (1.3)
  - Note: $\mathcal{C}_1 = \{ w \in T^+ \mid uw = v \text{ where } u, v \in \mathcal{C} \}.$
- For each $n \geq 1$; we then define $\mathcal{C}_\infty = \bigcup_{n=1}^{\infty} \mathcal{C}_n.$ (1.4)
  - Note: if $\mathcal{C}_{n-1} = \emptyset$ then $\mathcal{C}_n = \emptyset$,

Example 1.9: Compute C_n and C_∞

- Let C={0, 01, 011} as in Example 1.8. Then
- $C_1 =$? $C_2 =$? $C_n =$? for all $n \geq 2$ $C_\infty =$?

Algorithm to compute Cn

- Notation
  - Let A = “12”, B = “3xyz”, and C = AB, Then C = “123xyz”
  - A is a prefix of C and B is a postfix of C
  - Notation C/A denotes B

- Algorithm to compute C1, C2, …, Cn_1, Cn

```
C0 = C
For each code-word cw in C
    For each code cw_1 in Cn_1
        If cw_1 is prefix of cw, then add cw/cw_1 in Cn
        If cw is prefix of cw_1, then add cw_1/cw in Cn
```

The Sardinas-Patterson Theorem

- Theorem 1.10 (The Sardinas-Patterson Theorem)
  - A code C (finite) is uniquely decodable if and only if the sets C and C_∞ are disjoint. ($C \cap C_\infty = \emptyset$)
  - A code C (finite or infinite) is uniquely decodable if and only if $C \cap C_\infty = \emptyset$ and $C_n = \emptyset$ for some $n \geq 1$.

- Example 1.11
  - If $C = \{0, 01, 011\}$ as in Examples 1.8 and 1.9,
  - Then $\mathcal{C}_\infty = \{1, 11\}$ which is disjoint from C.

Example 1.12

- Let C be the ternary code {01, 1, 2, 210}.
  - Then C_1 = {10}, C_2 = {0} and C_3 = {1}, so $1 \in C \cap C_\infty$ and
  - thus C is not uniquely decodable.
- Can you find an example of non-unique decodability?