Coding Theory
and
Digital Data Transmission
Lesson X

Joel Isaacson
Bar-Ilan Computer Science

5th June 2005

1 Information Theory

Information theory provides a measure of the amount of the abstract concept of information that is contained within a data channel. This allows us to configure a channel that has sufficient capacity to carry the data traffic and to code the data in an optimal manner.

1.1 The Measure of Information

An information system is an object that produces **events**. Each event, x_i, occurs with a probability P_i. The events produces might be discrete or analog (continuous). A continuous source can be converted to a discrete source by sampling methods. A discrete information source either has a finite set of events that are possible or can be made finite, to a sufficient accuracy by a quantization step. The set of events can be mapped to elements of a set of symbols or letters.

Information sources can be classified as being either having **memory** or being **memoryless**. A source with memory will have the probability of the current symbol, s_i, dependent on previous symbols, s_{i-n}, where $n > 0$. Memoryless sources have each symbol not dependent on previous symbols.

The amount of information contained in an event is related to its **uncertainty**, or in the case of a memoryless information source the probability. The following criterion relate to the information content, $I(x_i)$, of each event i.

1. If the probability of an event is 1 the information contents is zero: $\{P(x_i) = 1\} \Rightarrow \{I(x_i) = 0\}$.
2. The information content is always positive: $\{I(x_i) > 0\}$.
3. Less probable events have greater information content: $\{P(x_i) < P(x_j)\} \Rightarrow \{I(x_i) > I(x_j)\}$.
4. Information is additive: $I(x_i x_j) = I(x_i) + I(x_j)$.

It is not difficult to show that the mathematical function the satisfies these relationships is:

$$I(x_i) = \log_b \frac{1}{P(x_i)} = -\log_b P(x_i). \tag{1}$$

The units of information is called **bits** if $b = 2$. This the the memory content of each event.

1.2 Average Information of an Event Source

In data communication we transmit a long string of symbols (events) from the information source. Thus we are interested in the average information content that the source produces. The mean value of $I(x_i)$ over the alphabet of the source with m different symbols is given by

$$H = \sum_{i=1}^{m} P(x_i)I(x_i) = -\sum_{i=1}^{m} P(x_i)\log_2 P(x_i) \quad \text{[b/symbol]}. \tag{2}$$

The quantity H is called from a rough analogy to a similar formula in statistical mechanics (Physics) the entropy of the data source. It is a measure of the average information content of the event source per character.

Note that for $m = 2^n$ events with equal probabilities we have n bits per symbol (H=$m\left(\frac{1}{m}\right)\log_2 m = \log_2 2^n = n$). In general, for any distribution of m events, the source entropy satisfies the following relationship

$$0 \le H \le \log_2 m. \tag{3}$$

The lower bound is for the case where one event has a probability of 1 (the other musts then have zero probability).

For a source that generates r symbols per second the information rate is

$$R = rH \quad \text{[b/s]}. \tag{4}$$

2 Source Coding

A conversion of the output of an event source into a sequence of binary symbols is called **source coding**. One objective of source encoding is to minimize the average bit rate required to represent the source by reducing the redundancy of the information source. The average code word length L,per symbol is given by

$$L = \sum_{i=1}^{m} P(x_i)n_i. \tag{5}$$

Where $n_i$is the number of bits for event x_i.

2.1 Source coding theorem

The source coding theorem states that the average code word length L, per symbol is bounded by the source entropy

$$L \ge H. \tag{6}$$

Of course since each concrete representation of an encoded symbol uses an integer number of bits we can not always approach the limit H.

2.2 Classification of codes

Classification of code is best illustrated by an example, table 1.

2.2.1 Fixed Length Codes

A fixed length code is one whose code word length is fixed. Code 1 and 2 of table 1 are fixed-length codes with length 2.

2.2.2 Variable Length Codes

A variable length code is one whose code word length is variable. Code 3, 4, 5 and 6 are variable length codes.

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

Table 1: Binary codes

2.2.3 Distinct codes

A code is distinct if each code word is distinguishable from the each other code. In table 1 all codes are distinct except for Code 1.

2.2.4 Prefix-free codes

A code in which no code word can be formed by adding one code symbol to another code word symbol is called prefix-free. Codes 2, 4 and 6 are prefix-free. Prefix-free codes are sometimes call **instantaneous codes** since they can be decoded as soon as the last bit is received.

2.2.5 Uniquely decodable codes

A code is uniquely decodable if the original code sequence can be reconstructed perfectly from the concatenation of any code sequence. All prefix-free codes are uniquely decodable. But some non prefix-free codes are uniquely decodable. Codes 2, 4, 5 and 6 are uniquely decodable.