Image Processing and Analysis

Chapter 8 Image Compression

Weiqiang Wang

School of Computer and Control Engineering, UCAS

December 31, 2015

Outline

Measuring Information

- The fundamental premise of information theory is that the generation of information can be modeled as a probabilistic process that can be measured in a manner that agrees with intuition.
- A random event E that occurs with probability $P(E)$ is said to contain $I(E)$ units of information.

$$I(E) = \log \frac{1}{P(E)} = -\log P(E)$$

The quantity $I(E)$ often is called the self-information of E.

Measuring Information(CONT.)

- The base of the logarithm in Eq. (1) determines the unit used to measure information. If the base m logarithm is used, the measurement is said to be in $m-ary$ units.

$$I(E) = \log \frac{1}{P(E)} = -\log P(E) \tag{1}$$

- If the base 2 is selected, the resulting unit of information is called a bit.

The Information Channel

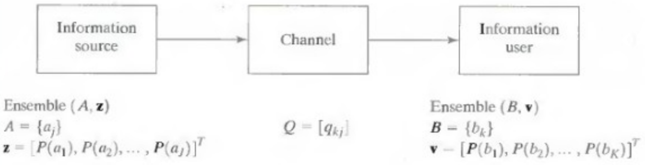

FIGURE 8.7 A simple information system.

- Assume that the information source in Fig generates a random sequence of symbols from a finite or countably infinite set of possible symbols. The set of source symbols $\{a_1, a_2, \ldots, a_J\}$ is referred to as the source alphabet A, and the elements of the set are called symbols or letters.

The Information Channel(CONT.)

- The probability of the event that the source will produce symbol a_j is $P(a_j)$

$$\sum_{j=1}^{J} P(a_j) = 1$$

- A $J \times 1$ vector $z = [P(a_1), P(a_2), \ldots, P(a_J)]^T$ customarily is used to represent the set of all source symbol probabilities $\{P(a_1), P(a_2), \ldots, P(a_J)\}$. The finite ensemble (A, z) describes the information source completely.

The Information Channel(CONT.)

- The self-information generated by the production of a single source symbol is $I(a_j) = -\log P(a_j)$.
- If k source symbols are generated, for a sufficiently large value of k, symbol a_j will (on average) be output $kP(a_j)$ times. Thus the average self-information obtained from k outputs is

$$-kP(a_1)\log P(a_1) - kP(a_2)\log P(a_2) - \ldots - kP(a_j)\log P(a_j)$$

- or

$$-k\sum_{j=1}^{J} P(a_j)\log P(a_j)$$

The Information Channel(CONT.)

- The average information per source output, denoted $H(z)$, is

$$H(z) = -k \sum_{j=1}^{J} P(a_j) \log P(a_j)$$

- This quantity is called the uncertainty or entropy of the source. It defines the average amount of information obtained by observing a single source output. As its magnitude increases, more uncertainty and thus more information is associated with the source. If the source symbols are equally probable, the entropy or uncertainty is maximized and the source provides the greatest possible average information per source symbol.

The Information Channel(CONT.)

- Because we modeled the input to the channel in $Fig.8.7$ as a discrete random variable, the information transferred to the output of the channel is also a discrete random variable($B : \{b_1, b_2, ..., b_K\}$).
- The probability of the event that symbol b_k is presented to the information user is $P(b_k)$. The finite ensemble (B, v), where $v = [P(b_1), P(b_2), ..., P(b_K)]^T$, describes the channel output completely and thus the information received by the user.
- The probability $P(b_k)$ of a given channel output and the probability distribution of the source z are related by the expression

$$P(b_k) = \sum_{j=1}^{J} P(b_k|a_j)P(a_j)$$

The Information Channel(CONT.)

- $P(b_k|a_j)$is the conditional probability that output symbol b_k is received, given that source symbol a_j was generated.
- The conditional probabilities are arranged in a matrix $K \times J$ matrix Q

$$Q = \begin{pmatrix} P(b_1|a_1) & P(b_1|a_2) & \dots & P(b_1|a_J) \\ P(b_2|a_1) & P(b_2|a_2) & \dots & P(b_2|a_J) \\ \vdots & \vdots & \vdots & \vdots \\ P(b_K|a_1) & P(b_K|a_2) & \dots & P(b_K|a_J) \end{pmatrix}$$

- The probability distribution of the complete output alphabet can be computed from

$$v = Qz$$

- Matrix Q, with elements $q_{kj} = P(b_k|a_j)$, is referred to as the forward channel transition matrix or by the abbreviated term channel matrix.

The Information Channel(CONT.)

- Each b_k has one conditional entropy function, this conditional entropy function, denoted $H(z|b_k)$, can be written as

$$H(z|b_k) = -\sum_{j=1}^{J} P(a_j|b_k) \log P(a_j|b_k)$$

- $P(a_j|b_k)$ is the probability that symbol a_j was transmitted by the source, given that the user received b_k. The expected (average) value of this expression over all b_k is

$$H(z|v) = \sum_{k=1}^{K} H(z|b_k) P(b_k)$$

The Information Channel(CONT.)

- After substitution and some minor rearrangement $H(z|v)$ can be written as

$$H(z|v) = -\sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j, b_k)\log P(a_j|b_k)$$

- $P(a_j, b_k)$ is the joint probability of a_j and b_k.That is, $P(a_j, b_k)$ is the probability that a_j is transmitted an b_k is received.
- The term $H(z|v)$ is called the equivocation of z with respect to v. It represents the average information of one source symbol, assuming observation of the output symbol that resulted from its generation.

The Information Channel(CONT.)

- Because $H(z)$ is the average information of one source symbol, assuming no knowledge of the resulting output symbol, the difference between $H(z)$ and $H(z|v)$ is the average information received upon observing a single output symbol. This difference, denoted $I(z,v)$ and called the mutual information of z and v, is

$$I(z,v) = H(z) - H(z|v)$$

- Substituting for $H(z)$ and $H(z|v)$, and recalling that $P(a_j) = P(a_j, b_1) + P(a_j, b_2) + ... + P(a_j, b_K)$, yields

$$I(z,v) = \sum_{j=1}^{J} \sum_{k=1}^{K} P(a_j, b_k) \log \frac{P(a_j, b_k)}{P(a_j)P(b_k)}$$

- After further manipulation, can be written as

$$I(z,v) = \sum_{j=1}^{J} \sum_{k=1}^{K} P(a_j) q_{kj} \log \frac{q_{kj}}{\sum_{i=1}^{J} P(a_j) q_{kj}}$$

The Information Channel(CONT.)

- The average information received upon observing a single output of the information channel is a function of the input or source symbol probability vector z and channel matrix Q.
- The minimum possible value of $I(z, v)$ is zero and occurs when the input and output symbols are statistically independent, in which case $P(a_j, b_k) = P(a_j)P(b_k)$.
- The maximum value of $I(z, v)$ over all possible choices of source probabilities in vector z is the capacity, C, of the channel described by channel matrix Q.That is,

$$C = \max_{z}[I(z, v)]$$

- The capacity of the channel defines the maximum rate at which information can be transmitted reliably through the channel. Moreover, the capacity of a channel does not depend on the input probabilities of the source but is a function of the conditional probabilities defining the channel alone.