XI. PROCESSING AND TRANSMISSION OF INFORMATION

Prof. S. H. Caldwell
Prof. P. Elias
Prof. R. M. Fano
Prof. D. A. Huffman
W. G. Breckenridge
J. Capon
E. Ferretti
E. J. McCluskey, Jr.
A. J. Osborne
R. A. Silverman
J. C. Stoddard
F. F. Tung
S. H. Unger
W. A. Youngblood

A. AN INFORMATION MEASURE FOR MARKOFF DIAGRAMS

A Markoff diagram is a linear graph whose nodes represent the states of a Markoff source and whose directed branches represent transitions between these states. These branches have associated with them transition probabilities. From these conditional probabilities the state probabilities themselves may be calculated from a set of linear equations. For instance, the diagram of Fig. XI-1 represents a four-state source of binary symbols. The heavily-lined branch tells us that, if the source is in state "4," the probability of the 0 symbol is 2/3 and it will be accompanied by a transition to state "1." If P_1, P_2, P_3, and P_4 are the state probabilities it follows from the diagram that

$$\begin{cases} P_1 = \frac{3}{4} P_3 + \frac{2}{3} P_4 \\ P_2 = \frac{1}{3} P_1 + \frac{1}{2} P_2 + \frac{1}{3} P_4 \\ P_3 = \frac{2}{3} P_1 \\ P_4 = \frac{1}{2} P_2 + \frac{1}{4} P_3 \end{cases}$$

These equations may be solved for

$$P_1 = \frac{12}{43}, \; P_2 = \frac{14}{43}, \; P_3 = \frac{8}{43}, \text{ and } P_4 = \frac{9}{43}$$

From a knowledge of the symbol source an information measure may be associated with a sequence of symbols from the source. This calculation may be made symbol by symbol. For example, if we know the source to be in state "4," the information carried by the symbol 0 will be $-\log_2 \frac{2}{3}$ bits. The average information/symbol of our example may be calculated to be

$$\begin{aligned} H_{av}\text{ (sequence)} &= P_1\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right) + P_2\left(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}\right) \\ &\quad + P_3\left(-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}\right) + P_4\left(-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}\right) \\ &= 0.925 \text{ bits/symbol} \end{aligned}$$

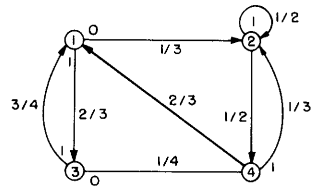

Fig. XI-1
A Markoff diagram.

It is as reasonable to compute an information measure for the Markoff diagram itself as it is to attach an information measure to the sequences from a Markoff source. To understand this viewpoint, let us imagine that the diagram is an air view of a highway system with intersections and one-way roads. Suppose that the getaway car of a bank robber has escaped into this maze of roads. A pursuing highway patrol car desires to follow in the tire tracks of the escaping car.

Consider the problem that confronts the patrolman when he reaches an intersection where a sign informs him: "The escape car followed the road marked "0" with the probability one-half, or the road marked "1" with the probability one-half." (See state "2" of Fig. XI-1.) The chase cannot continue with assurance until one bit of information becomes available. A knowledge of the actual road followed would give just this amount. The road sign itself furnishes no information.

Now the chase leads past a sign reading: "Escape cars take the left road with a probability two-thirds and the right road with a probability one-third." (See state "4" of Fig. XI-1.) Knowledge of the escape route would, at this location, furnish on the average an information of $-\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.919$ bits. (This corresponds to the information content of a single symbol of the sequence.) The difference between this and the one bit necessary for the decision to be made by the patrol car is $1.000 - 0.919 = 0.081$ bits. Thus it is convenient to say that this quantity is the average information given by the road sign.

The maximum possible information would be given by a sign which read: "Escape cars always turn left here." This would furnish one bit of information to a patrolman who, previous to reading the sign, assumed that the two roads were equiprobable for escape.

In all of these cases the average information furnished by the road signs added to the average information given by a knowledge of the exact escape road (the latter information based on the probabilities listed on the sign) is just one bit. Thus, by analogy, the average information that may be associated with a Markoff diagram for a binary source is

$$H_{av\ (diagram)} = 1 - H_{av\ (sequence)} \text{ bits/transition}$$

The diagram of Fig. XI-1 has, consequently, an information measure of

$$1.000 - 0.925 = 0.075 \text{ bits/transition}$$

In general, a Markoff diagram for an n-symbol source has a measure of

$$H_{av\ (diagram)} = \log_2 n - H_{av\ (sequence)} \text{ bits/transition}$$

and the information output rate of a first-order Markoff source and the information measure of the diagram which describes the source are complementary to each other.

D. A. Huffman

B. ON BINARY CHANNELS AND THEIR CASCADES

The following material is a summary of an investigation which will be reported in greater detail in the future. This condensation does not contain the detailed derivations nor the visual aids that a full understanding of the problem requires.

1. Capacity and Symbol Distributions

Many interesting features of binary channels are concealed if only symmetric channels are considered, as is usually the case. Accordingly, we have applied Muroga's formalism (1) to a detailed discussion of the arbitrary binary channel.

Let the channel have a matrix

$$\begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} \tag{1}$$

Then the capacity is

$$c(\alpha, \beta) = \frac{-\beta H(\alpha) + \alpha H(\beta)}{\beta-\alpha} + \log\left[1 + \exp_2\left(\frac{H(\alpha) - H(\beta)}{\beta-\alpha}\right)\right] \qquad \exp_2 x \equiv 2^x \tag{2}$$

The probability that a zero is transmitted if capacity is achieved is

$$P_o(\alpha, \beta) = \beta(\beta-\alpha)^{-1} - (\beta-\alpha)^{-1}\left[1 + \exp_2\left(\frac{H(\beta) - H(\alpha)}{\beta-\alpha}\right)\right]^{-1} \tag{3}$$

and the probability that a zero is received if the signaling is at capacity is

$$P_o'(\alpha, \beta) = \left[1 + \exp_2\left(\frac{H(\beta) - H(\alpha)}{\beta-\alpha}\right)\right]^{-1} \tag{4}$$

$H(x)$ is the entropy function $-x \log x - (1-x) \log (1-x)$.

Each of the functions in Eqs. 2, 3, and 4 defines a surface over the square $0 \leqslant \alpha \leqslant 1$, $0 \leqslant \beta \leqslant 1$, with certain symmetries intimately related to the noiseless recodings possible if the designation of input symbols or output symbols or both is changed. It is found that an infinite number of binary channels have the same capacity. Each equicapacity line has two branches, one above the $\beta = \alpha$ line, the other below it. If C is a channel on one branch, then so is ICI, but IC and CI are on the other branch. By I we mean the matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The $P_0(\alpha, \beta)$ surface is saddle-shaped, and has discontinuities at the corners $\alpha=\beta=0$ and $\alpha=\beta=1$. Any limiting value between $1/e$ and $1-(1/e)$ can be achieved by approaching these points in the proper direction. In particular, if the point $\alpha=\beta=0$ is approached along the positive α-axis, the limit is $1/e$; along the positive β-axis, the limit is $1-(1/e)$. These two limits are the maximally asymmetric input symbol distributions for a binary channel; that is, no input symbol distribution more asymmetric than $[1/e,\ 1-(1/e)]$ or $[1-(1/e),\ 1/e]$ can achieve capacity in a binary channel. The low-capacity channels which in the limit of vanishing ϵ exhibit this maximum asymmetry are the channels

$$C_\epsilon = \begin{bmatrix} \epsilon & 1-\epsilon \\ 0 & 1 \end{bmatrix}, \ IC_\epsilon,\ C_\epsilon I, \text{ and } IC_\epsilon I \tag{5}$$

2. Probability of Error

The probability of error at capacity of the channel (Eq. 1) is

$$P_e(\alpha, \beta) = \beta + (1-\alpha-\beta)\ P_0(\alpha, \beta)$$

and is the same as that of the channel ICI, and one minus that of IC and CI. It follows that if P_e is greater than 1/2 (which it is, if the channel is on the upper branch of the equicapacity line) then P_e can be made less than 1/2 by the simple expedient of reversing the designation of the input (or output) symbols. Of all the channels on the lower branch of an equicapacity line, the symmetric channel has the smallest probability of error, unless the capacity is low, in which case some asymmetric channel has the smallest probability of error.

3. Cascaded Channels

Cascading a binary channel corresponds to squaring its channel matrix. It is found that any channel on the upper branch of an equicapacity line, except the symmetric channel, has a lower capacity in cascade than its images on the lower branch. Thus, even if no delay can be tolerated in a cascade of channels, the intermediate stations should cross-connect the outputs of a preceding stage to the inputs of the next stage if the channel lies on the upper branch of an equicapacity line. It can be shown that this behavior is minimum probability of error detection, provided rate-destroying mappings are precluded.

Of the channels on the lower branch of the equicapacity line, the symmetric channel has the highest capacity in cascade, unless the capacity is low, in which case some asymmetric channel has the highest capacity in cascade.

4. Redundancy Coding in ϵ-Channels

If sufficient delay is permitted at intermediate stations, it follows from Shannon's second coding theorem that the end-to-end capacity of a cascade of identical channels can be made arbitrarily close to the common capacity of the separate channels. Simple redundancy coding is quite effective (though not ideal) for the low-capacity ϵ-channels (Eq. 5), and serves to illustrate how delay can be exchanged for enhanced rate in a cascade of channels. The coding consists of repeating each digit r times at the transmitter and (synchronously) decoding the received output in blocks of r digits: a run of r ones is identified as a one, and a sequence of r digits with a zero at any position is identified as a zero, since in the channel C_ϵ a zero at the receiver means that a zero was transmitted. Simple though this coding scheme is, it gives rates that are an appreciable fraction of capacity with a very small probability of error. Consequently, the capacity of a long cascade of redundancy-coded ϵ-channels can be kept much higher (at the cost of sufficient delay) than if no such coding were permitted.

R. A. Silverman

References

1. S. Muroga, On the capacity of a discrete channel, I, J. Phys. Soc. Japan 8, 4, 484-494 (1953).

C. MEASUREMENTS OF STATISTICS OF PICTURES

Analog measurements of the second-order probability distributions of several pictures were made. The procedure and equipment used were briefly discussed in the Quarterly Progress Report, April 15, 1955, page 49. A value of comentropy for each picture was calculated from the measured probabilities. Noise measurements from the equipment permitted bounds to be placed on the value of comentropy obtained.

Two-dimensional autocorrelation measurements on the same pictures were made and a second value of entropy obtained for each picture from these statistical measurements. The equipment used for both measurements, as well as the results obtained, is described fully in Bounds to the Entropy of Television Signals, Jack Capon, M. Sc. Thesis, Department of Electrical Engineering, M.I.T., 1955. A technical report based on parts of this thesis will be published.

The digital equipment described in the April report, page 52, was checked and put into operation. Complete second-order probability measurements on a picture were made. The equipment is still rather critical in operation, but improvements should permit more reliable measurements.

The equipment and results of tests, as well as the second-order probabilities of a picture, are described in Measurement of Second-Order Probability Distribution of Pictures by Digital Means, J. C. Stoddard, M. Sc. Thesis, Department of Electrical Engineering, M.I.T., 1955. A technical report based on parts of this thesis will also be published.

J. C. Stoddard