THE GAME OF "GOSSIP" ANALYZED BY THE THEORY OF INFORMATION

Edgar Reich

P-201

15 March 1951

The RAND Corporation · SANTA MONICA · CALIFORNIA

ABSTRACT

These paper deals with an analysis of a distributed model of the game of "gossip", in which a message is passed through a line of individuals, and the final (in general, garbled) result is compared with the original ungarbled message. The deterioration of information (defined in the sense of Shannon and Wiener) along the line is calculated, and exact as well as asymptotic formulas suggest approximate linear electric network analogues. ( )

I. Object of this Paper

In discussing communication nets involving human beings and/or other automata one is tempted to call upon analogies from the theory of linear electric network to describe the mutual exchange of information on terms of analogous current or voltage propagation. In this paper we consider a particular type of communication system, and investigate the approximations under which for that particular case an analogy can rigorously be drawn, using the strict definition of information proposed by Shannon and Wiener (1)

II. The Theory of Information

Consider a communication system (Fig. 1) described by the square nxn matrix

$P(i,j)$ = joint probability that the i^{th} symbol is sent, and, as a result of this, the j^{th} symbol received. (1)

---

(1) Confer list of references at end of paper.

Define

$P(i)=\sum_{j=1}^{n} P(i,j)$ = a-priori probability of sending symbol i. (2)

$Q(j)=\sum_{i=1}^{n} P(i,j)$ = a-priori probability of receiving symbol j. (3)

$Q_i(j)=P(i,j)/P(i)$ = conditional probability that j is (4)

recieved if i is sent.

It is evidently necessary that

$Q_i(j)=1, \quad (i=1,2,\ldots,n).$ (5)

Shannon has shown [2] that the information rate of the communication system of Fig. 1 should be defined as

$R=H-E$, where (6)

$H=-\sum_{j=1}^{n} Q(j)\log Q(j)$ is the negentropy of the received symbols[3] (7)

$E=-\sum_{i,j=1}^{n} P(i,j)\log Q_i(j)$ is the equivocation of the system. (8)

In the special case when all the symbols to be sent are equally likely we have

$P(i)=1/n.$ (9)

Therefore, $P(i,j)=(1/n)Q_i(j)$, giving

$E=-(1/n)\sum_{i,j=1}^{n} Q_i(j)\log Q_i(j).$ (10)

---

(2) Cf. ref. 7, pg. 409; also see ref. 5 for a discussion of the arbitrariness of the choice of the definition of information. Shannon uses $H(y)$ for our H, and $H_x(y)$ for our E. Shannon uses logs to the base 2; we employ natural logs.

(3) The term "negentropy" is due to Brillouin, ref. 2.

Also, if in addition, $Q_i(j)=Q_j(i)$

$$Q(j)=\sum_{i=1}^{n} P(i)\, Q_i(j)=(1/n)\sum_{j=1}^{n} Q_i(j)=(1/n) \text{ by (5).} \tag{11}$$

In this particular case, therefore, H=logn, so that

$$R=\log n+(1/n)\sum_{i,j=1}^{n} Q_i(j)\log Q_i(j). \tag{12}$$

III. Model for the Game of Gossip

The game is postulated to consist in selecting a symbol (message) at random from a set of n symbols, such that all n symbols are a-priori equally likely, and propagating it down the line of individuals. The participants are assumed to know what the a-priori possible messages are, and to realize that they are a-priori equally likely.

Let x measure distance from the beginning, or origin of the line.

We now make the idealizing postulate that if the symbol arriving at x is the i^{th} symbol, then the probability that an arbitrary symbol j arrives at x+dx is given by the conditional probability matrix

$$\left[Q_i(j)_{x,x+dx}\right]=\begin{bmatrix} 1-\alpha(n-1)dx & \alpha dx & \cdots & \alpha dx \\ \alpha dx & 1-\alpha(n-1)dx & \cdots & \alpha dx \\ \alpha dx & \alpha dx & & 1-\alpha(n-1)dx \end{bmatrix}$$

Here α should be thought of as a number proportional to the density of human beings per unit distance along the line. In the realistic case distance would be measured as proportional to the number of human beings. Therefore α plays the role of a characteristic frailty (independent of n) which must be found by experiment [4].

Symbolically

$$Q_i(j)_{x,x+dx}=(1-andx)\ \delta_{ij}+adx^{(5)}. \tag{13}$$

Let

$f_{ij}(x)$=conditional probability that j will be received at x (14)

if i is inserted at x=0.

Then, using (13), (5)

$$f_{ij}(x+dx)=\sum_{k=1}^{n} f_{ik}(x)\ Q_k(j)_{x,x+dx}=f_{ij}(x)-f_{ij}(x)\ andx+adx \tag{15}$$

Thus $f_{ij}(x)$ satisfies the differential equation

$$f'_{ij}\ (x)+\ anf_{ij}(x)-a=0 \tag{16}$$

The solution, with the initial condition $f_{ij}(0)=\delta_{ij}$ is

$$f_{ij}(x)=e^{-anx}\delta_{ij}+\ (1-e^{-anx})/n. \tag{17}$$

IV. Information Flow in the Game of Gossip

According to (10), if E(x) is the equivocation in the stretch of line (0.x),

$$E(x)=\ -(1/n)\sum_{i,j=1}^{n} f_{ij}(x)\log f_{ij}(x). \tag{18}$$

---

(4) Although an exact solution for the case of discrete individuals could have been obtained the continous case was purposely chosen because it is simpler.

(5) δ_{ij} is the Kronecker delta: $\delta_{ij}=\begin{cases}1 \text{ if } i=j\\ 0 \text{ if } i\neq j\end{cases}.$

Substituting (17), manipulating algebraically, and inserting into (12) gives as the basic result

$$R_n(x) = \frac{1+(n-1)e^{-\alpha nx}}{n} \log\left[1+(n-1)e^{-\alpha nx}\right] + \frac{(n-1)}{n}(1-e^{-\alpha nx})\log(1-e^{-\alpha nx}). \tag{19}$$

Here $R_n(x)$ is the information available at x regarding the original message. A plot of the function is given in Fig. 2. As obvious checks we have

$$\lim_{x\to 0} R_n(x) = \log n \tag{20}$$

$$\lim_{x\to\infty} R_n(x) = 0 \tag{21}$$

$$R_1(x) = 0 \tag{22}$$

$$\lim_{n\to\infty} R_n(x) = 0 \quad \text{when } x \neq 0. \tag{23}$$

It is interesting to evaluate the derivative of $R_n(x)$. This is

$$R_n'(x) = \alpha(n-1)\, e^{-\alpha nx} \log\left[\frac{1-e^{-\alpha nx}}{1+(n-1)\, e^{-\alpha nx}}\right] \tag{24}$$

This shows that

(a) $R_n(x)$ is strictly decreasing as a function of x except when $\alpha(n-1) = 0$ in which case $R_n(x) \equiv$ const.

(b) The slope of $R_n(x)$ at the origin is infinite when $\alpha(n-1) \neq 0$.

V. Estimates and Asymptotic Relations

The following two relations are easily obtained from (19):

For small x:

$$R_n(x) = \log n + \alpha (n-1)\, x\log x + O(x^2 \log x). \tag{25}$$

For large n, and αx larger than some positive quantity independent of n:

$$R_n(x) = (3n/2)e^{-2\alpha n x} + O(e^{-2\alpha n x}). \tag{26}$$

An interesting relation that can also be obtained is

$$\lim_{n\to\infty} \frac{R_n\left[\frac{1}{\alpha(n-1)}\right]}{R_n(0)} = 1/e \doteq 0.37. \tag{27}$$

This indicates th t for large n if we go out a distance $1/[\alpha(n-1)]$ from the origin we will find that the information has deteriorated to approximately 37% of its initial value.

VI. Linear Electric Network Analogies

Linear electric net-work analogy approximations can be obtained by exploiting either (26) or (27). (25) cannot be used because the xlogx term prevents it from being the solution of a linear differential equation with constant coefficients.

An electric analogy could use either time or distance as the quantity analogous to x. If distance is used, we have the case of an electric transmission line with attenuation. Recognizing that a series combination of resistance R and capacitance C has a time constant RC seconds (time in which the original capacitor voltage goes to $1/e \approx 0.37$ of its initial value) a simple lumped analogy making use of (27) can be obtained. Consider namely a capacitance of value 1 farad and leakage resistance $[1/\alpha(n-1)]$ ohms (Fig. 3). If the capacitance is originally charged to log n volts then the capacitor voltage as a function of time is approximately analogous to the information in the gossip line as a function of distance.

VII. Supplementary Remark

Of the several possible modifications of the preceding analysis that come to mind the most obvious one is the adoption of a different $_1(j)_{x,x+dx}$. This will simply change the differential equation (16), but the rest of the work would be the same. One possible example is to make α a function of x. This would reflect the effect of relative position in the group on the frailty of the individual.

REFERENCES

1. L. Brillouin, "Life, Thermodynamics, and Cybernetics", Am. Sci., 37, 4, (Oct. 1949), 554-568.

2. L. Brillouin, "Thermodynamics and Information Theory", Am. Sci., 38, 4, (Oct. 1950), 594-599.

3. N. Rashevsky, "Some Bio-Sociological Aspects of the Mathematical Theory of Communication", Bull. Math. Bioph., 12, 4, (Dec. 1950) 359-378.

4. R.C. Raymond, "Communication, Entropy, and Life", Am. Sci., 38, 2, (April 1950), 273-278.

5. E. Reich, "Definition of Information", Procs. Inst. Rad. Engrs., 39, 3, (March 1951), 290.

6. S.O. Rice, "Communication in the Presence of Noise", Bell System Tech. J., 29, 1, (Jan. 1950), 60-93.

7. C.E. Shannon, "A Mathematical Theory of Communication", Bell System Tech. J., 27, 3, (July 1948), 379-423.

8. C.E. Shannon, "A Mathematical Theory of Communication (Concluded)", Bell System Tech. J., 27, 4, (Oct. 1948), 623-656.

9. C.E. Shannon, "Communication in the Presence of Noise", Procs. Inst. Rad. Engrs., 37, 1, (Jan. 1949), 10-21.

10. C.E. Shannon, "Communication Theory of Secrecy Systems", Bell System Tech. J., 28, 4, (Oct. 1949), 656-715.

11. C.E. Shannon, "Prediction and Entropy of Printed English", Bell System Tech. J., 30,1, (Jan. 1951), 50-64.

12. N. Wiener, "Cybernetics", John Wiley, New York, 1948, pp. 74-81.

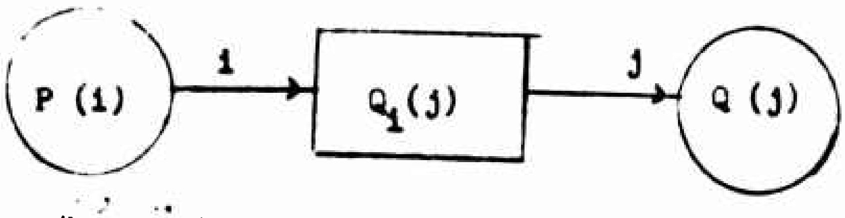

Fig. 1. Fundamental Communication system

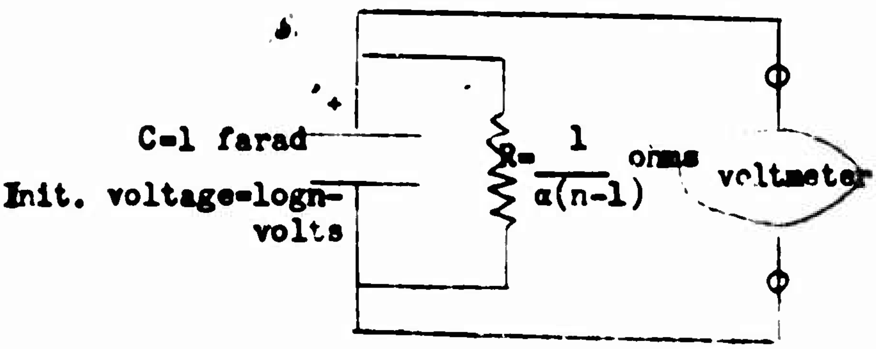

Fig. 2 Approximate anology for gossip line having frailty per unit distance α, for case of n possible messages. Here time is the quantity analogous to distance along the gossip line, and voltage is analogous to information.

Fig. 3 Information along gossip line as a function of distance from origin.