A Text Recognition Procedure for Cryptanalysis

BY C. V. KIMBALL

Unclassified

Presented before the International Symposium in Information Theory at UCLA, 31 January–2 February 1966, this paper reveals the character and quality of work in cryptanalysis now in progress in universities.

It uses standard Bayesian methods equivalent to the common log scoring techinques with which our analysts are familiar. The conditions for acceptance or rejection of the null hypothesis are somewhat more explicitly stated, perhaps, but no new concepts are involved. The solution given is for the easy part of the problem, with all probabilities assumed known, and even a priori odds. The really challenging task is to deduce useful weights, not from known or assumed probabilities, but from samples of the cipher or code. The problem treated is clearly and carefully stated, and the solution is thorough and competent.

Dr. B. C. Getchell, P1

INTRODUCTION

This paper is a brief summary of a long-term study reported in detail in [1]. As a result, much of the underlying material is given only cursory treatment. The first section introduces the decision procedure used in the study and explains the importance of source redundancy in making rapid decisions. In the second section, the decision procedure is applied to the recognition of natural-language texts among random texts. In the third section, the recognition procedure is applied to a general cryptographic problem.

DEFERRED-DECISION PROCEDURE

In the binary detection problem considered here, one of two sources, *SN* (signal plus noise) or *N* (noise alone), provides a stream of *M* symbols to the decision device.[1] The decision device is to respond *A* if *SN* is present, or *B* if *N* is present. Decisions are made on the basis of sequential observations of successive symbols from the source. The two sources generate independent symbols taken from the same finite alphabet, $(Z_1, Z_2, \ldots, Z_K)$, according to known probability distribu-

[1] Much of the current work in deferred-decision theory involves the detection of signals in noise, and the symbols above are common.

tions which are conditional to the source. We let α_i be the probability of Z_i when SN is present and let β_i be the probability of Z_i when N is present. In addition, we will assume that the *a priori* probabilities of SN and N are known.

For this problem, the deferred-decision procedure leads to minimum expected losses according to a predetermined loss rule. This loss rule assigns a cost W_M to the response B when the SN source is present, and a cost W_F to the response A when N is present. Also, a fixed cost, C, is charged for each observation. When the parameters M, $\{\alpha_i\}$, $\{\beta_i\}$, W_F, W_M, and C are specified, the decision procedure can be found by using a computer-implemented optimization process.

The log-odds-ratio transformation is helpful in describing the operation of the decision device. Let m designate the number of observations that have been taken; then the log-odds-ratio after m observations is given by

$$L_m = ln \frac{P(SN|m \text{ observations})}{1 - P(SN|m \text{ observations})} . \tag{1}$$

When the state of the decision process is expressed in terms of L_m, Bayes' rule can be written in a form particularly convenient for analysis and implementation:

$$L_{m+1} = L_m + \lambda(Z_i). \tag{2}$$

Here L_{m+1} is the log odds ratio after the symobl Z_i is observed, and λ_i is the log likelihood ratio of Z_i, given by

$$\lambda (Z_i) = ln \alpha_i/\beta_i. \tag{3}$$

The decision process is usually considered in terms of L_m. The decision function is represented by $M+1$ pairs, (Δ_m, Γ_m), of decision points. If $L_m \geq \Delta_m$, the decision is A; if $L_m < \Gamma_m$, the decision is B. If $\Delta_m \geq L_m > \Gamma_m$, another observation is taken and the process continues. Fig. 1 is helpful in visualizing the operation of the device.

As background, let us consider an important theorem for the case in which the symbols from the N source have a uniform distribution —that is, in which $\beta_i = 1/K$ for all i. For sequential procedures, as is well known, ([2], [3]), the speed with which decisions are made depends on $\Delta\mu$, the difference in mean values of the log likelihood ratio in SN and N.

$$\Delta\mu = E(\lambda \,|SN) - E(\lambda \,|N). \tag{4}$$

The following theorem relates $\Delta\mu$ to the Shannon-redundancy of the SN source.

Theorem

$$\Delta\mu \geq ln2 \left[\log_2 K + \sum_{i=1}^{i=K} \alpha_i \log_2 \alpha_i \right] \tag{5}$$

$$\geq ln2 \, R(SN) . \tag{6}$$

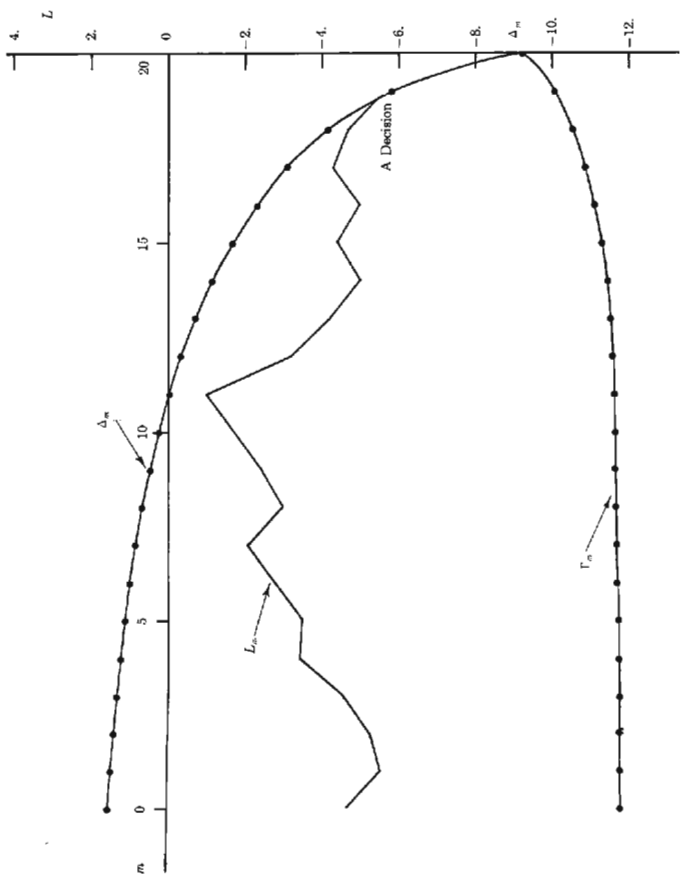

Fig. 1—Operation of the decision procedure.

Thus one can often compare the detectability of two sources when only their respective redundancies are known.

THE RECOGNITION PROCEDURE

The problem considered here is that of recognizing natural-language among random texts in which symbols are uniformly distributed. Natural-language text is defined as a meaningful sequence of symbols from a written language—that is, letters spelling words which, together, convey a meaning. A deferred-decision procedure is used to recognize the natural language rapidly and achieve optimum loss performance.

The deferred-decision procedure can be used only when distributions of the symbols for random and natural-language texts differ. The theorem in Section 1 suggests that the redundancy of natural-language text is an excellent measure of the differences between the two distributions. Since the single-symbol redundancy of most natural languages is of the order of 15 percent, the deferred-decision procedure provides an effective recognition technique. For purposes of analysis, we have considered the problem of recognition of English text; of course, the procedure can be applied to other written languages.

The deferred-decision procedure requires that the loss parameters W_F, W_m, and C, be specified; they, in turn, depend on the problem at hand. Here we will use the loss conditions that arise in the application of the procedure to cryptography, described in the next section. In this application the loss structure has the following characteristics.

(1) The cost of a miss, W_m, is much greater than that of a false alarm, W_F.

(2) The cost of observation for a single symbol is much smaller than that of either kind of error loss.

The present analysis is based on the loss ratio $W_m/W_F = 500.0$ and $W_F/C = 10{,}000$, and yields representative results for this loss structure.

The deferred-decision procedure was analyzed for the above loss conditions for the case in which 36 observations were available to the decision device; that is, $M = 36$. The probabilities α_i were taken as the single-letter probabilities for English and are shown with their corresponding λ's in Table 1. Figure 2 depicts the decision functions obtained and the motion of the log-odds ratio for two texts. The first text is from a well known news magazine, the second has been derived with the aid of a table of random units. Decision procedure uses *a priori* probability of English of 0.000002. The theoretical performance of the procedure under these conditions is given as a function of the probability of SN in Table 2, where ν_{SN} and ν_N designate

the expected number of observations in SN and N, respectively. From these results we conclude that English text can be recognized very rapidly by the deferred-decision procedure.

Symbol, Z_i	α_i	β_i	$\lambda(Z_i)$
A	0.815	0.0385	+0.7510
B	0.0144	0.0385	−0.9817
C	0.0276	0.0385	−0.3322
D	0.0379	0.0385	−0.0149
E	0.1311	0.0385	+1.2258
F	0.0292	0.0385	−0.2737
G	0.0119	0.0385	−0.6564
H	0.0526	0.0385	+0.3131
I	0.0635	0.0385	+0.5008
J	0.0013	0.0385	−3.3720
K	0.0042	0.0385	−2.2145
L	0.0339	0.0385	−0.1262
M	0.0254	0.0385	−0.4161
N	0.0710	0.0385	+0.6129
O	0.0800	0.0385	+0.7316
P	0.0198	0.0385	−0.6624
Q	0.0012	0.0385	−3.4590
R	0.0683	0.0385	+0.5747
S	0.0610	0.0385	+0.4616
T	0.1047	0.0385	+1.0011
U	0.0246	0.0385	−0.4469
V	0.0092	0.0385	−1.4315
W	0.0154	0.0385	−0.9153
X	0.0017	0.0385	−3.1428
Y	0.0198	0.0385	−0.6624
Z	0.0008	0.0385	−3.9110

Table 1.—Symbol probabilities for English and random texts.

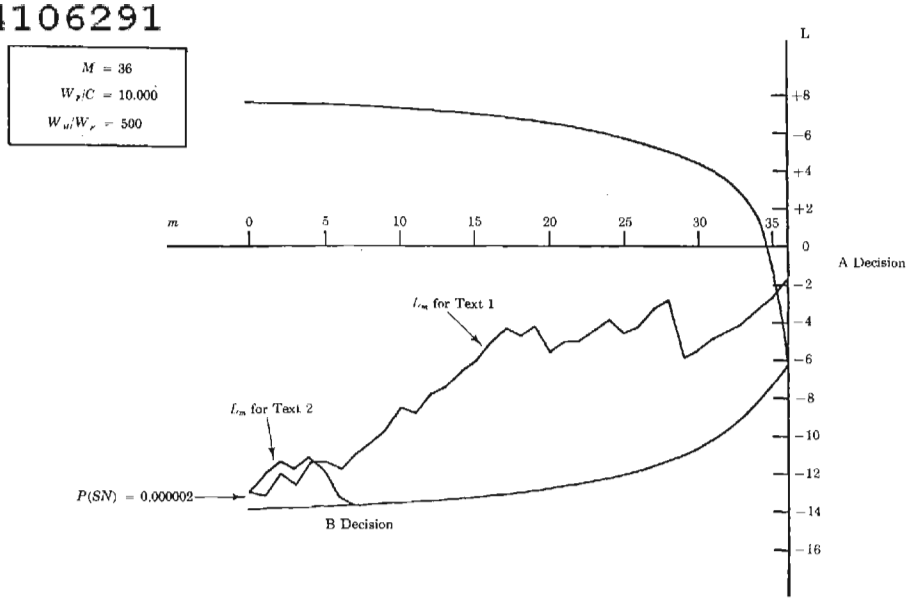

Fig. 2.—Decision points and operation of the recognition procedure.

Text 1: LEGED CONNECTION TO CIVIL RIGHTS X IN SHORT

Text 2: TRCSG VJMUY GFADF AYHBA NZNIV MIGEX MDHVZ T

A Priori Probability Of SN	*Probability of Correct Detection*	*Probability of False Alarm*	$^{\nu}SN$	$^{\nu}N$
0.000002	0.754	<0.0001	27.	3.3
0.00001	0.873	0.0002	33.	6.1
0.0001	0.983	0.0006	33.	10.0
0.001	0.996	0.0021	31.	13.0
0.01	0.999	0.0065	27.	17.0
0.1	0.999	0.0174	24.	21.0

Table 2.—Theoretical performance of the recognition procedure.

$$W_F/C = 10.000 \qquad W_M/W_F = 500 \qquad M = 36$$

APPLICATION TO CRYPTANALYSIS

The recognition procedure described in the preceding section was developed for use in cryptanalysis, the extraction of meaningful text from an enciphered text without a key. Although a particular cipher is used as an example, the method presented here is applicable to any cipher that preserves the redundancy of the concealed text.

A cipher is a set of invertible transformations $\{f_j\}$ of a natural language text t into an enciphered text x,

$$f_j(t) = x. \tag{7}$$

The subscript j designates the particular transformation being used and is referred to as the key. The basic problem of cryptanalysis is to determine f_j and t from a given x. Shannon has shown [4] that cryptanalysis can be successful only if the concealed text contains redundancy.

A fundamental approach to cryptanalysis is to consider the set of all possible inverses of the message x, $\{f_j^{-1}(x)\}$. If x has a large enough number of characters (greater than what Shannon calls the unicity distance), there will be a most probable text t^* in $\{f_j^{-1}(x)\}$. This most probable text t^* will have the same statistical structure as the natural language. Thus an enciphered text can be analyzed by examining all possible inverses $f_j^{-1}(x)$ for the structure of the natural-language text.

Since the effectiveness of the recognition procedure depends on the redundancy of the natural language, the procedure is well suited for the above approach. In addition, the performance of the recognition procedure can be predicted directly from the theory.

Let us now apply the recognition procedure to an example of cryptanalysis. A cipher using 456,976 possible keys was used to en-

cipher a typical segment of English text. This cipher, suggested by Shannon ([4], p. 709), produces enciphered texts with nearly uniform single-letter distributions. A computer was programmed to test all possible keys on the enciphered text, using the recognition procedure. The time per trial of a 72-character message was less than 100 seconds.

Table 3 compares the experimental results with those predicted by the theory. Since the maximum number of observations for the experiment, 72, was twice the number used in the theoretical analysis, the results for the experiment should be slightly better than those of the theory; and, indeed, except for the ν_{SN}, the experimental results are in excellent agreement with the theory. The difference in ν_{SN} is due to the differences in the maximum number of observations.

	Theoretical ($M=36$)	*Experimental* ($M=72$)
Probability of Correction Detection	0.754	0.8
Probability of False Alarm	<0.0001	0.00008
ν_{SN}	27.	70.[2]
ν_N	3.3	3.5

Table 3.—Comparison of the theoretical and experimental results.

$W_F/C = 100$ $\quad W_M/W_F = 500$
A Priori Probability of $SN = 0.000002$

CONCLUSIONS

The speed of detection for a deferred-decision process has been related to the source redundancy when the N source has uniform distribution. This relation has been used to develop a recognition procedure for natural-language text. The theoretical performance of the recognition procedure has been supported by results obtained in the solution of a general problem of cryptanalysis.

[2]Discrepancy between theoretical and experimental results is accounted for by the increased M of the experimental work.

REFERENCES

[1] C. V. Kimball, *A Recognition Procedure for Natural-Language Text with Application to Cryptography*, Unpublished thesis, Department of Electrical Engineering, The University of Michigan, Ann Arbor, 1965.
[2] A. Wald, *Sequential Analysis*, John Wiley and Sons, Inc., New York, 1947.
[3] T. G. Birdsall and R. A. Roberts, *Theory of Signal Detectability: II: Decision Processes*, Cooley Electronics Laboratory Technical Report No. 136, The University of Michigan, Ann Arbor, 1964.
[4] C. E. Shannon, "Communications Theory of Secrecy Systems," *Bell Systems Technical Jourrnal*, Vol. 28, No. 4, October 1949, p. 656.