APPLICATION OF SIMULATION TO
DESIGN OF CLASSIFICATION SYSTEMS

Robert DeMoyer, Jr.
Dr. Jon I. Fellers

General Electric Company
Philadelphia, Pennsylvania

Abstract

In the future, a dramatic increase is likely in use of digital computers to make routine but very complex decisions, particularly decisions involving quantified uncertainty and large amounts of data. This paper formulates a method to classify objects on the basis of a series of imperfect measurements, the results of which may be either discrete or continuous. Digital simulation is demonstrated to be a powerful tool to effect quantitative trades among factors important in the implementation of the classification problem. Basic ideas are illustrated using a contrived example of inference of a satellite mission.

Introduction

As computer oriented mathematical techniques become more routinely applied to the solution of engineering problems, implementation of systems to solve the problem of classification will become more common. This problem, as defined for this discussion, is the important application of decision theory represented by Figure 1.

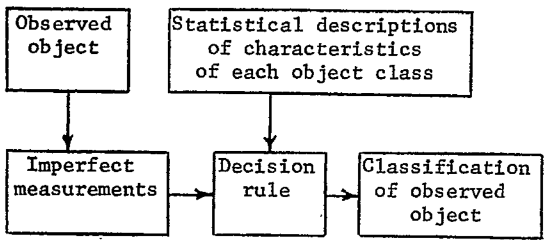

Figure 1.

The Classification Problem

Suppose a given set of object classes are each characterized by a set of parameters. These parameters do not take on unique values for each object class, but are represented by conditional probability density functions. The classification problem is then given a set of imperfect measurements of the characteristic parameters of an object, to decide to which class the object belongs.

A simple example of the classification problem is the detection of slugs in a vending machine. There are two object classes, slugs and legitimate coins. One might choose, as a set of characteristic parameters, weight, diameter, and presence of magnetic material. Clearly, these parameters will be described by different (prior) density functions for each object class. Engineers have attempted to design machines which solve this classification problem with a high degree of accuracy.

In this paper, we are concerned with the design of complex systems to solve more complicated classification problems. In implementing such a system, the engineer generally has a choice of many parameters to observe, what decision rule is to be followed, and with what accuracy to perform the measurements. Also, economic considerations tend to restrict the choice of observed parameters and measurement accuracy. Digital simulation will be shown to be a powerful tool which permits quantitative trades among factors important in the implementation of solutions of the classification problem.

Important practical applications of the classification problem include the classification of diseases based on several symptoms, or the identification of the mission of an artificial satellite based on certain of its characteristics. In the field of disease classification, Warner[1], et al., has used a fixed set of tests to classify congenital heart disease, and Gorry and Barnett[2] have advanced a method of sequential testing. In neither case is measurement error considered, nor is simulation employed to evaluate the accuracy of the system.

To solidify the ideas developed in this paper, a contrived example of the determination of the mission of a man-made satellite will be presented.

Analysis

Prior Data

Suppose a situation exists where there are N classes of objects, and M tests, the results of which are to be used in classification. For each of the M tests, there is one density function for each of the N object classes. This body of M•N density functions is designated by $P_i\ (s|V_j)$. For example, suppose that Test 1 is the measurement of length. If there are four possible classes, $P_i\ (s|V_j)$ might be as shown in Figure 2. In this figure, the letter s refers to a state, which is a range of values. In this case, the range of each state is five feet. For object class number one, the density function is read as follows:

$$P_1(s=2|V_1) = .1; \quad P_1(s=3|V_1) = .6; \quad P_1(s=4|V_1) = .3$$

In many problems, determination of acceptable prior density functions is a difficult task. The density function of the diameter of a true coin could be approximated by measuring a large number of coins. The variance for the diameter of a slug is likely greater because of the large variety of slugs. In this case, some judgement as to how a slug might be made is useful in construction of the prior data. Finally, the determination of the density functions of a satellite vehicle would have to depend almost entirely on engineering judgment as to how it might be constructed. If there is a way to observe the success and failure of the decision process, these prior density functions may be modified as more is learned about the nature of the various classes. In this respect, a system could be devised which has a capacity for learning.

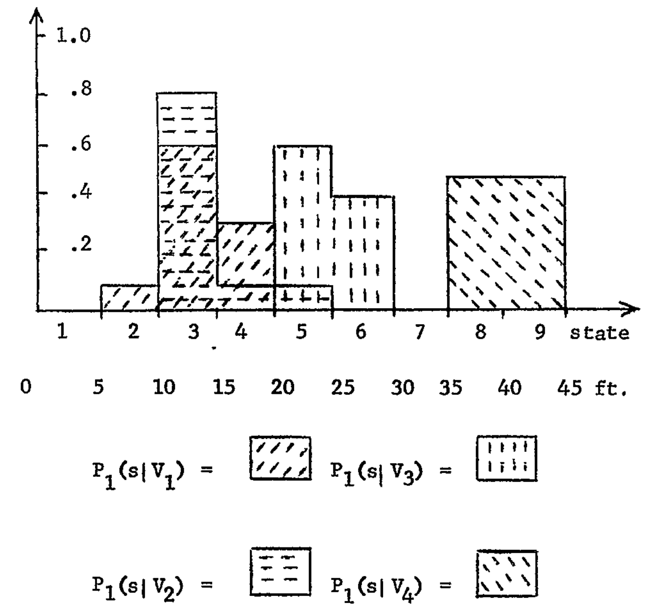

Figure 2.

Length Prior Distribution

Decision Process

The results of each of M measurements are contained in the vector $\underline{r}$. Each measurement, regardless of its dimension, fits into a predetermined range, called a state. Therefore, for every $\underline{r}$ there is a unique state $\underline{s}$, but not vice versa. If V_j refers to the j^{th} object class, the following is a conditional probability:

$$P(V_j|\underline{s}) = \frac{P(\underline{s},V_j)}{P(\underline{s})} = \frac{P(\underline{s}|V_j)\ P(V_j)}{P(\underline{s})} \qquad (1)$$

where

$P(V_j|\underline{s})$ = the probability of V_j being present such that the measurement $\underline{r}$ falls into state $\underline{s}$.

$P(\underline{s}, V_j)$ = the probability of the measurements being in state $\underline{s}$, and of V_j being present.

$P(\underline{s})$ is the total probability of the measurements falling into state $\underline{s}$.

$P(V_j)$ = the prior probability of class V_j being present.

These values are determined from prior knowledge of the frequency of occurrence of the object classes. Having calculated this, the <u>decision rule</u> is to pick vehicle class (i) if

$$P(V_i|\underline{s}) > P(V_j|\underline{s}) \qquad \forall \quad j \neq i \qquad (2a)$$

or,

$$P(\underline{s}|V_i)P(V_i) > P(\underline{s}|V_j)P(V_j) \qquad \forall \quad j \neq i \qquad (2b)$$

where

$P(\underline{s}|V_j)$ = the probability that the measurements fall into $\underline{s}$ when V_j is present.

Note that the denominator $P(\underline{s})$ of Equation (1) does not appear in Equation (2b) because $P(\underline{s})$ has the same value for all j.

The former is calculated as follows:

$$P(\underline{s}|V_j) = P_1(s_1|V_j)P_2(s_2|V_j)\ldots.P_M(s_M|V_j) \qquad (3)$$

where $P_i(s_i|V_j)$ is the probability that the state measured in the i^{th} test is s_i when the class V_j is present. The product is taken under the assumption of independence of tests. The decision rule adopted is a special case of Bayes decision rule where the costs of all wrong decisions are equal, and the cost of a correct decision is zero.

A discussion now is required to show how the values are determined for the terms of Equation (3). For the moment it will be assumed that there is no measurement error. Suppose that the first test is length and that the density functions are those shown in Figure 2. If the measurement is r_1 = 18 ft., then it lies in state $s_1 = 4$. If there are four possible classes, Equation (3) is evaluated for j = 1, 2, 3 and 4. The first term of each results from this length measurement. From Figure 2,

$$\begin{array}{ll} P_1(s_1{=}4|V_1) = 0.3 & P_1(s_1{=}4|V_3) = 0. \\ P_1(s_1{=}4|V_2) = 0.1 & P_1(s_1{=}4|V_4) = 0. \end{array} \qquad (4)$$

It is obvious that $P(\underline{s}|V_3) = P(\underline{s}|V_4) = 0$, and that only $P(\underline{s}|V_1)$ and $P(\underline{s}|V_2)$ need be considered from this point on. The next test supplies values for $P_2(s_2|V_1)$ and $P_2(s_2|V_2)$. The procedure continues through all M tests, then the decision is made by Equation (2b).

In discussing measurement error, we will replace Equation (3) by the following:

$$P(\underline{r}|V_j) = P_1(r_1|V_j)P_2(r_2|V_j)\ldots.P_M(r_M|V_j) \qquad (5)$$

This is done to emphasize the fact that r_i, which is the i^{th} measurement, does not necessarily fall into the state in which the true dimension will fall. In order to make a decision, the terms $P_i(r_i|V_j)$ are needed for Equation (5), but, unlike $P_i(s_i|V_j)$, these cannot be obtained directly from the prior density functions.

As previously mentioned, a test result may be continuous such as length. On the other hand, a test may be discrete. For example, the material in a slug may be steel, brass, lead or plastic. In this case, a measurement is an entire state, not a value within a state. The performance of a measuring device for discrete tests may be specified by a matrix of probabilities $P(r{=}\ell|s{=}k)$ having such terms as the probability of measuring a slug as brass when it is truly steel. This matrix, designated by E_i, is shown in Figure 3. Note that as measurement error approaches zero, E_i becomes the identity matrix.

$$\begin{array}{l} \\ \text{True state } s{=}1 \\ \text{True state } s{=}2 \\ \vdots \end{array} \begin{array}{cc} \text{Meas. } r{=}1 & \text{Meas. } r{=}2 \\ \left[\begin{array}{c} P(r{=}1|s{=}1) \\ P(r{=}1|s{=}2) \\ \vdots \end{array}\right. & \left.\begin{array}{cc} P(r{=}2|s{=}1) & \ldots \\ P(r{=}2|s{=}2) & \ldots \\ \vdots & \end{array}\right] \end{array} = E_i$$

Figure 3.

Discrete Error Specification

The terms of Equation (4) corresponding to discrete tests are given by the following total probability:

$$P_i(r_i|V_j) = \sum_{k=1}^{KT_i} P(r_i|s=k)P_i(s=k|V_j) \qquad (6)$$

where KT_i = number of states in test i. The first term is contained in the matrix E_i, and the second term is contained in the prior probability density functions for the i^{th} test.

The situation for continuous tests is not quite so straight forward. While $P(r_i|s=k)$ was conveniently available for discrete tests, two problems exist for continuous tests. First, $P(r_i|s=k)$ is actually zero because r_i represents a particular value, not a state. To give it a value, the probability must be of r_i falling into some range of values. A convenient range is the state into which it falls. Secondly, if the true state is state k, the true value is assumed equally likely to be anywhere within that state. For that reason, $P(r_i=\ell|s=k)$ will mean the average probability of the measure falling into the state ℓ, containing the measurement r_i, as the true value ranges across state k.

In order to calculate this expected probability, the nature of the continuous error must be specified. It was found computationally convenient to specify this error by a tolerance E, such that if the true value is γ, the measured value will fall anywhere between $\gamma(1-E)$ and $\gamma(1+E)$ with equal probability.

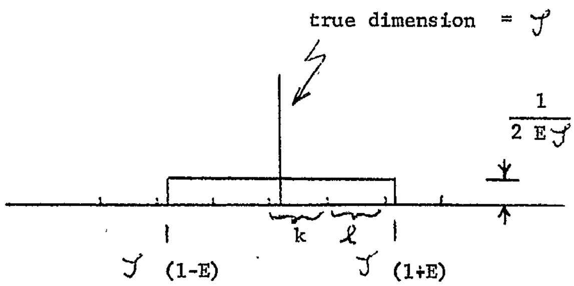

Figure 4.

Continuous Error Specification

For a specific dimension γ, contained in state k, the probability that the measure will fall into state ℓ is shown in Figure 4 to be $1/2E\gamma$ times the range of values in state ℓ, given by $R_{\ell+1} - R_\ell$.

$$P(\ell|\gamma \in k) = \frac{R_{\ell+1} - R_\ell}{2E\gamma} \qquad (7)$$

γ can be anywhere in state k, so the expected probability that a measurement will fall into state ℓ, given that the true state is k, is the following:

$$P(r_i=\ell|s=k) = \int_{R_k}^{R_{k+1}} \frac{R_{\ell+1} - R_\ell}{2E\gamma} \frac{d\gamma}{R_{k+1} - R_k} \qquad (8)$$

This integral must be carried out for all states k to provide values for Equation (6). The various combinations of integration limits in Equation (8) will not be noted here.

Referring to Figure 4, as $E \to 0$, the only state into which the measurement can fall is the state containing the true dimension. After correctly evaluating the boundary conditions,

$$\lim_{E \to 0} P(r_i=\ell|s=k) = \delta_{\ell,k} \qquad (9)$$

This is analagous to the discrete error matrix approaching the identity matrix. Note that when $P(r_i=\ell|s=k) = \delta_{\ell,k}$, then Equation (6) reduces to the no error case.

$$P_i(r_i|V_j) \xrightarrow[\text{error} \to 0]{} P_i(s_i|V_j) \qquad (10)$$

Without the foregoing treatment of measurement error, the error will cause some measurements to be fit into wrong states, increasing classification error. The method used takes advantage of the known measurement error statistics to reduce misclassification. In the limit when error causes a test to be completely inconclusive, the outcome of the test will have no effect on the classification decision.

Simulation

At this point, the decision rule has been established and the equations needed to handle observed data have been developed. For the purpose of evaluating the effectiveness of a system, large numbers of "designs" of various classes are made by Monte Carlo on the basis of the density functions corresponding to the class to be simulated. Next, the design is "measured" by a random number chosen as a function of the design dimension and the known error statistics of each test. These measurements are then fed into the classification process, and the resulting decision is compared to the class which was simulated. The results for a large number of runs are displayed by a matrix P where P_{ij} approaches the probability of deciding that the j^{th} class is present when the i^{th} class is in fact present. This simulation process is illustrated in Figure 5. When the tests employed are quite distinctive and the errors of the measuring devices small, the matrix P will be quite diagonal. It is often the case that a single figure of merit is desired for an entire system in order to make meaningful comparisons of various system implementations. In this case, a good one is the product of the diagonal terms of P.

There are many uses for the results of simulation. First of all, it shows how well a completely specified system can be expected to perform. It is often the case that the number of tests to be performed is less than the total number that could be. Simulation using various combinations of tests can lead to selection of an optimal combination. Also, the accuracy with which measurements are made may be a variable. In this case, the simulation results provide data for trades among total number and combination of tests, and the accuracy of the measurement of the tests.

Whenever a Monte Carlo simulation is made, it is of interest to know how many runs constitute a sufficient number of runs. To get an approximate feel for this, a specific example, which will be detailed in the next section, was taken A particular class was simulated thirty times at one hundred runs each. An approximate application of the central limit theorem showed that a simulation of one class using 1,000 runs will result in P_{ij} being within 0.012 of the true value with

90% confidence.

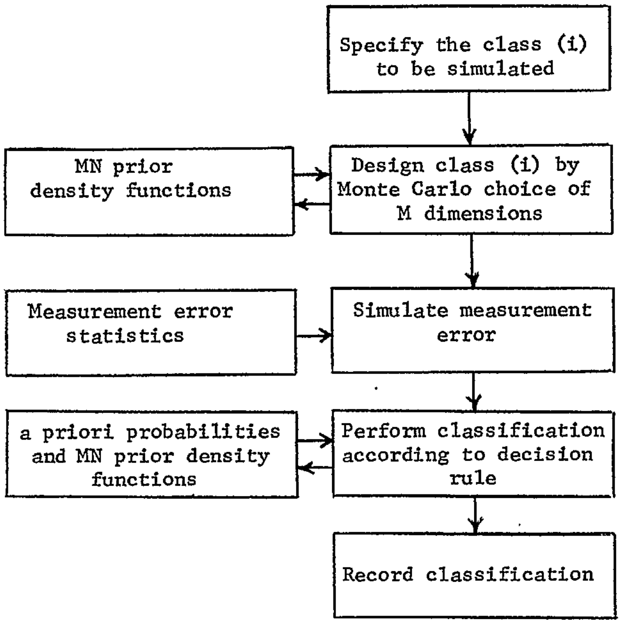

Figure 5.

Simulation Procedure

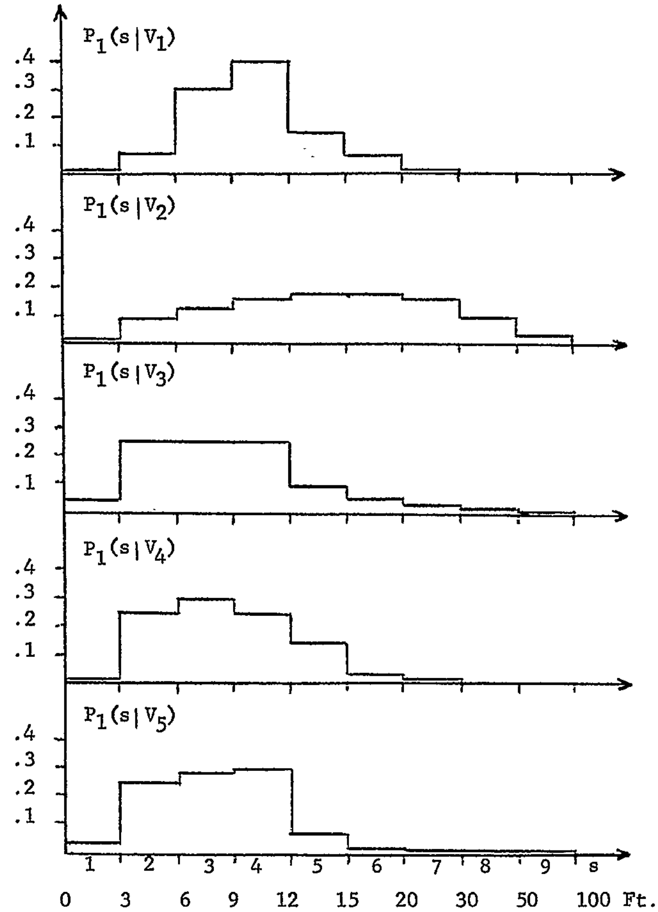

Figure 6.

Length Distributions For The Example

Example

To illustrate the concepts in this paper, a simple example has been contrived. For the purpose of the example, all possible satellite missions are assumed to fall into five classes (N=5):

1. Orbital Weapon
2. Low Altitude Observation
3. Communications
4. Meteorological
5. General Scientific

The mission characteristics are inferred from various sensors (M=10):

1. Length
*2. Heat Shield Presence
3. Velocity Change Capability
4. Average Power Level
*5. Communications Mode
*6. Environmental Control Method
7. Perigee Altitude
8. Orbit Inclination
9. Argument of Perigee
10. Number of Similar Vehicles in Orbit

* discrete tests

The prior distributions for Length are shown in Figure 6. The remaining distributions are not shown due to space limitations. Two observation systems are considered:

1. Co-orbital observation of satellite
2. Ground based observation only

All vehicle classes are assumed equally likely to appear, so $P(V_j) = 1/5$, $j = 1,5$. The discrete tests, marked (*), have discrete error matrices E_i, which become more diagonal as error decreases (see Table 1).

Note first that the ground based sensors have no ability to measure heat shield presence or environmental control mode. Also, number of similar vehicles in orbit is actually a discrete measure, but is approximated as a continuous test due to the fact that the discrete error matrix could be 50 x 50 or larger.

Remembering that P_{ij} is the probability of deciding class j is present when class i is, the P matrices resulting from 500 simulations of each class for each error level are as follows:

$$P_1 = \frac{1}{500} \begin{bmatrix} 487 & 5 & 3 & 0 & 5 \\ 6 & 433 & 4 & 29 & 28 \\ 1 & 5 & 435 & 32 & 27 \\ 0 & 23 & 20 & 402 & 55 \\ 3 & 38 & 23 & 71 & 365 \end{bmatrix}$$

$$P_2 = \frac{1}{500} \begin{bmatrix} 457 & 27 & 7 & 0 & 9 \\ 9 & 417 & 11 & 29 & 34 \\ 5 & 7 & 405 & 47 & 36 \\ 1 & 32 & 31 & 376 & 60 \\ 7 & 42 & 28 & 76 & 347 \end{bmatrix}$$

TEST		COORBITAL	GROUND
1. Length		$E = 0.02$	$E = 0.5$
2. Heat Shield	present absent	$\begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} = E_2$	$\begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix} = E_2$
3. Velocity Change		$E = 0.5$	$E = 0.99$
4. Average Power		$E = 0.3$	$E = 0.99$
5. Communications Mode	None Beacon Telemetry Command Video	$\begin{bmatrix} .8 & .05 & .05 & .05 & .05 \\ .05 & .8 & .05 & .05 & .05 \\ .05 & .05 & .8 & .05 & .05 \\ .05 & .05 & .05 & .8 & .05 \\ .05 & .05 & .05 & .05 & .8 \end{bmatrix} = E_5$	$\begin{bmatrix} .6 & .1 & .1 & .1 & .1 \\ .1 & .6 & .1 & .1 & .1 \\ .1 & .1 & .6 & .1 & .1 \\ .1 & .1 & .1 & .6 & .1 \\ .1 & .1 & .1 & .1 & .6 \end{bmatrix} = E_5$
6. Environmental Control	None Passive Temperature Control Active Temperature Control Heat Exchanger Controlled Atmosphere	$\begin{bmatrix} .7 & .08 & .08 & .07 & .07 \\ .08 & .7 & .08 & .07 & .07 \\ .07 & .08 & .7 & .08 & .07 \\ .07 & .07 & .08 & .7 & .08 \\ .07 & .07 & .08 & .08 & .7 \end{bmatrix} = E_6$	$\begin{bmatrix} .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \\ .2 & .2 & .2 & .2 & .2 \end{bmatrix} = E_6$
7. Perigee Altitude		$E = .01$	$E = .01$
8. Orbital Inclination		$E = .01$	$E = .01$
9. Argument of Perigee		$E = .01$	$E = .01$
10. Number of Similar Vehicles in Orbit		$E = .5$	$E = .75$

Table 1

Measurement Error Specification

Note that the accuracy of classification is greater for the case of lesser error. There are many ways of evaluating these results, depending on the emphasis of false alarm, false dismissal, or correct detection. A single figure of merit is the diagonal product which is the probability of correctly detecting one of each class.

$$DP_1 = 0.537$$

$$DP_2 = 0.403$$

Summary

In the future, a dramatic increase is likely in use of computers to make routine but very complex decisions, particularly decisions involving quantified uncertainty and large amounts of data. This paper has formulated the classification problem in a form suitable for solution by digital computers. Furthermore, the utility of digital simulation in implementing systems which solve the classification problem is demonstrated.

1 "Experience with Bayes Theorem for Computer Diagnosis of Congenital Heart Disease", Homer R. Warner, Alan F. Toronto and L. George Veasy, Annals of the New York Academy of Sciences 115, pp. 558-567, 1964.

2 "Experience with a Model of Sequential Diagnosis", G. Anthony Gorry and G. Octo Barnett, Computers and Biomedical Research, pp. 490-507, 1968.