STATISTICAL PATTERN RECOGNITION AND LEARNING FOR CONSTRUCTION ROBOTS

Kingsley Harrop-Williams[1]

INTRODUCTION

The ability to recognize, remember and classify patterns is the primary function of the brain. These are done so readily by humans that there is a tendency to take them for granted. However, this ability is the basic definition of intelligence. An "intelligent" construction robot would seem unearthly if it does not possess this ability. In this paper a decision function and learning algorithm is outlined for construction robots to recognize, classify and remember basic features of objects within their work environment.

Construction robots are foreseen in building construction, agriculture, mining, undersea, space, etc. Hence, unlike manufacturing robots, they must work with unstructured tasks in unstructured environments. As with the environment in which the construction robot must work, its data, be it visual, audio, chemical or radar, is random. In dealing with this type of data the recognition principle to use must be based on statistical decision theory. This allows the distinction between objects based on the mean value and scatter of their respective sensor values and treats any deterministic phenomenon (zero scatter) as a special case.

The method adopted is to reduce a large n-dimensional random vector obtained from sensor data to a much smaller m-dimensional transformed vector which retains all the statistical information of the data. Using this transformed vector (or the orginial vector) a statistical decision is made that optimally distinguishes between different objects. The accuracy of the parameters of the decision function is sharpened recursively by a Bayesian learning procedure such that even in an unknown environment the robot can gradually learn the objects around it.

PROBABILISTIC PATTERNS IN SENSOR INFORMATION

Construction robots gather information about their surroundings by means of sensors. In order to distinguish one object from another one generally compare the magnitude of the sensor data obtained from the two objects. As a simple example, consider the metal detector. Here the magnitude of the data from metalic objects are much larger than their surroundings.

---

[1]Assistant Professor, Dept. of Civil Engineering, Carnegie-Mellon University, Pittsburgh, PA 15213

If one discretize the sensor information as shown in figure 1, then the sensed value x can be used to identify the object from which it was taken. However, if the frequency of x is plotted

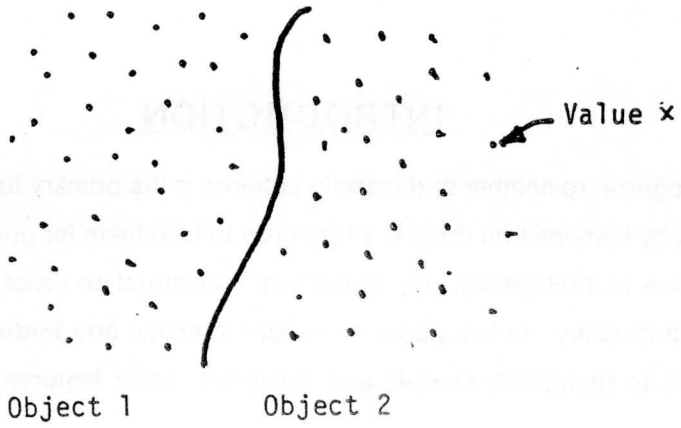

Figure 1 Discretized Sensor Information

against the magnitude of x one obtains the patterns shown in figure 2. Basically, the sensor values of object 1 forms a cluster around its mean value M_1, and those object 2 around M_2. The spread of the frequency distributions around the mean values is the result of the randomness of nature. Well-defined objects produce little spread, while ill-defined objects produce spreads around the mean values that may overlap.

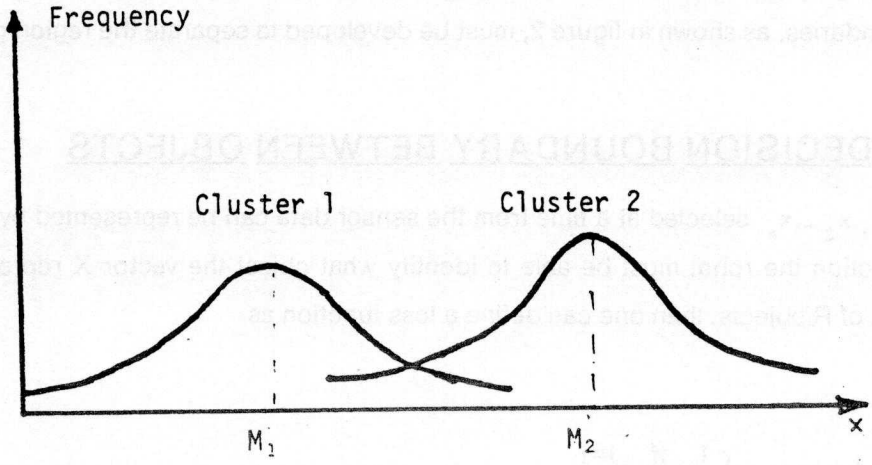

Figure 2 Frequency of sensor data

Selecting two values x_1 and x_2 at a time, rather than one from figure 1 a two-dimensional frequency plot can be constructed. This is illustrated in figure 3, with the frequency axis perpendicular to the paper. This again indicates clusters around M_1 for object 1 and around M_2 for object 2, with some overlap of the spread around the mean values.

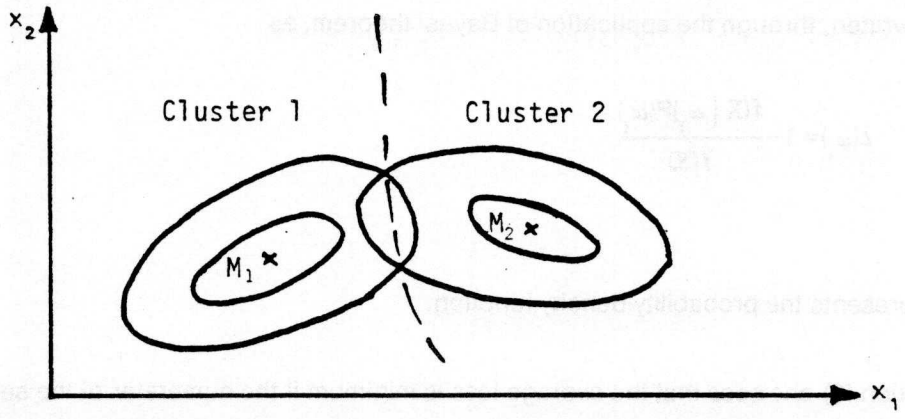

Figure 3 Frequency for two values

For a speedy analysis, the selection of one or two values at a time from the sensor data is inappropriate. The selection of some large number (n) of values must be made. The result is n-dimensional frequency distributions representing each object. These distributions invariably overlap and decision boundaries, as shown in figure 2, must be developed to separate the region governed by each object.

DECISION BOUNDARY BETWEEN OBJECTS

The n values $x_1, x_2, \ldots, x_n$ selected at a time from the sensor data can be represented by a vector X. As part of its function the robot must be able to identify what object the vector X represents. If ω_i represents the ith of R objects, then one can define a loss function as

$$L(\omega_i \mid \omega_j) = 1 - \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

That is, a unit loss when a wrong decision is made and zero loss if the decision is correct. The average loss as determined from equation (1) is

$$L(\omega_i) = \sum_j L(\omega_i \mid \omega_j) P(\omega_j \mid \mathbf{X}) = 1 - P(\omega_i \mid \mathbf{X}) \tag{2}$$

where $P(\omega_i|X)$ is the posterior probability that the object is ω_i given the vector X. Equation (2) can further be rewritten, through the application of Bayes' theorem, as

$$L(\omega_i) = 1 - \frac{f(\mathbf{X} \mid \omega_i) P(\omega_i)}{f(\mathbf{X})} \tag{3}$$

where f( ) represents the probability density function.

From equation (3) one sees that the average loss is minimum if the numerator of the second term on the right hand side is maximum. As is done conventionally, the decision function can be taken as the logarithm of this term, or

$$d_i(\mathbf{X}) = \ln f(\mathbf{X} \mid \omega_i) + \ln P(\omega_i) \tag{4}$$

If a construction robot has within its memory the frequencies $f(\mathbf{X} \mid \omega_i)$ and probabilities $P(\omega_i)$ of R objects expected to be encountered in its working environment, then a sampled vector X of the measured sensor data is identified as belonging to the object with the largest $d_i(X)$. The error associated with this identification is the sum of the error of omitting ω_i and the error of including ω_j, or

$$\varepsilon = P(\omega_i \mid X \epsilon \omega_j) + P(\omega_j \mid X \epsilon \omega_i) \text{ for all } j \neq i \tag{5}$$

In the decision function identified above, the robot memory consists of the frequency function $f(X \mid \omega_i)$ for each object ω_i. This frequency function can be predetermined in the laboratory. If the identified function can be approximated by the normal distribution it can be written as

$$f(\mathbf{X} \mid \omega_i) = \frac{1}{(2\pi)^{n/2} \mid C_i \mid^{1/2}} exp\left\{ -\frac{1}{2}(\mathbf{X} - M_i)^T C_i^{-1} (\mathbf{X} - M_i) \right\} \tag{6}$$

where M_i and C_i are the mean vector and covariance matrix of object ω_i. In this case the decision function becomes

$$d_i(\mathbf{X}) = ln P(\omega_i) - \frac{n}{2} ln(2\pi) - \frac{1}{2} ln \mid C_i \mid - \frac{1}{2}[(\mathbf{X} - M_i)^T C_i^{-1} (\mathbf{X} - M_i)] \tag{7}$$

Since the second term is the same for all ω_i it can be neglected and equation (7) is rewritten as

$$d_i(\mathbf{X}) = b_i - \frac{1}{2}[(\mathbf{X} - M_i)^T C_i^{-1} (\mathbf{X} - M_i)]$$

where

$$b_i = ln P(\omega_i) - \frac{1}{2} ln \mid C_i \mid \tag{8}$$

Equation (8) shows that the decision function is a guadratic in X. For two objects it is as illustrated in figure 3.

STATISTICAL LEARNING

The pattern recognition problem can be viewed as that of generating the decision boundaries which separates the R objects on the basis of the measured sensor data. The robot memory consists of the frequency distributions representative of each object. For the normal distribution and most common distributions, this consists of only two terms, the mean vector M_i and covariance matrix C_i. The general situation is that the true values of M_i and C_i are never available and the memory is first designed based on a small labelled set of samples. However, as exploration progresses and the robot identifies objects these parameters can be made to approach optimum or satisfactory forms.

To illustrate the technique of statistical learning, let $\underline{\theta} = [M_i, C_i]$ be the parameter vector. The distribution of X can be written in terms of $\underline{\theta}$ as

$$f(\underline{X} \mid \underline{\theta}) = \sum_{j=1}^{R} f(\underline{X} \mid \omega_j, \underline{\theta}_j) P(\omega_j) \tag{9}$$

Also, our knowledge about $\underline{\theta}$ is from a set of K independent sample vectors

$$\chi_k = (\underline{X}_1, \underline{X}_2 \ldots \underline{X}_k) \tag{10}$$

with frequency distribution $f(\underline{\theta} \mid \chi_k)$ easily determined from sample statistics. If another sample vector X_{k+1} is obtained then, from Bayes' theorem

$$f(\underline{\theta} \mid \chi_{k+1}) = f(\chi_{k+1} \mid \underline{\theta}) f(\underline{\theta}) / \int f(\chi_{k+1} \mid \underline{\theta}) f(\underline{\theta}) d\underline{\theta} \tag{11}$$

From independence of samples

$$f(\chi_{k+1} \mid \underline{\theta}) = \prod_{h=1}^{n} f(\chi_h \mid \underline{\theta}) = f(\chi_{k+1} \mid \underline{\theta}) f(\chi_k \mid \underline{\theta}) \tag{12}$$

which substituted into (11) gives the Bayes' learning algorithm

$$f(\underline{\theta} \mid \chi_{k+1}) = f(\chi_{k+1} \mid \underline{\theta}) f(\underline{\theta} \mid \chi_k) / \int f(\chi_{k+1} \mid \underline{\theta}) f(\underline{\theta} \mid \chi_k) d\underline{\theta} \tag{13}$$

Equation (13) shows that the new memory of the robot $f(\underline{\theta} \mid \chi_{k+1})$ is a modification of the old memory $f(\underline{\theta} \mid \chi_k)$ by the information contained in the newly sampled vector X_{k+1}. This updated parameter vector $\underline{\theta}$ updates M_i and C_i which in turn sharpen the decision boundary $d_i(X)$.

FEATURE SELECTION

The number of values n selected from the sensor data is arbitrary. However, for speed of processing the information collected by the sensors it may be necessary that this number be quite large. This results in large and cumbersome frequency distributions. Since most of those values contain the same information, it may be advantageous to select only those with dominant features and construct the decision function based on the features alone. The idea is illustrated in figure 4.

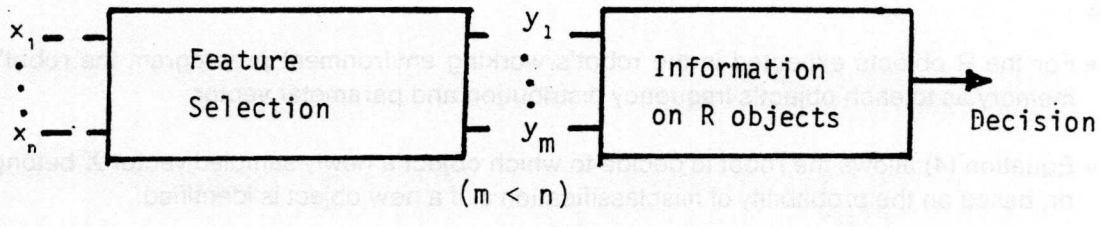

Figure 4. Recognition scheme

One way to identify features contained in the random vector X is by means of the Karhunen-Loeve expansion (Fukunaga, 1972). Here X is transformed into a new random vector Y with elements (i = 1 ...n)

$$y_i = \Phi_i^T X \tag{14}$$

where Φ_i is the i^{th} eigenvector of the covariance matrix of X. Each component of y is a feature that contributes to X.

The features y_i are mutually uncorrelated and their variances are given by the corresponding eigenvalue λ_i. A measure of the information contained in feature y_i is its variance λ_i. Hence dominant features have larger variances. If a feature is neglected it can be shown (Fukunaga, 1972) that the mean-square error introduced is the variance of that feature. It follows that if the eigenvalues are indexed such as $\lambda_1 > \lambda_2 > \ldots > \lambda_n$, then the n-dimensional problem $x_1, x_2, \ldots, x_n$ can be reduced to one in m dimensions ($n > m$) $y_1, \ldots y_m$ with a mean-square error of

$$\epsilon^2_{(m)} = \sum_{i=m+1}^{n} \lambda_i \tag{15}$$

SUMMARY AND DISCUSSION

The procedure described above for robot recognition and learning can be outlined by the following steps.

- For the R objects expected in the robot's working environment preprogram the robot's memory as to each object's frequency distribution and parameter vector.
- Equation (4) allows the robot to decide to which object a newly sampled vector X belong; or, based on the probability of misclassification ε, if a new object is identified.
- If a sampled vector X is identified with an acceptable probability of misclassification, as belonging to a certain object, Equation (13) allows the robot to sharpen its memory about that object.

For large amount of sensor information, working with the originial data may be slow and cumbersome. In this case a feature selection technique is outlined that allows an n-dimensional problem to be reduced to a m-dimensional one ($m<n$). From this point the steps outlined above can be followed using the m-dimensional feature vector. Other methods of feature selection can be found in MacQueen (1967) and Fukunaga and Koontz (1970).

The decision function described here is Bayesian, and provides optimum separation of the objects. However, the probability density function and probability concerning the objects are needed as input. Although sufficent laboratory investigation can establish these quantities, other decision functions like best-fit linear, piece-wise linear or polynomial can be developed (Tou and Gonzalez, 1974).

REFERENCES

1. Fukunaga, K. (1972) "Introduction to statistical pattern recognition", Academic Press, N.Y.

2. Fukunaga, K. and Koontz, W.L. (1970) "Representations of random processes using the finite Karhunen-Loeve expansion" Information and control 16, 83-101.

3. MacQueen, J. (1967) "Some methods for classification and analysis of multivariate observation" 5th Proc. Berkeley Symp. on Probability and Statistics, pp. 281-297.

4. Tou, J.T. and Gonzalez, R.L. (1974) "Pattern recognition Principles", Addison-Wesley, London.