1979 Carnahan Conference on Crime Countermeasures
University of Kentucky, Lexington, Kentucky — May 16-18, 1979

CATASTROPHE THEORY AND SOME DESIGN ASPECTS OF AUTOMATED IDENTIFICATION SYSTEMS

by

David McGoveran
SRI International
Menlo Park, California 94025

Abstract. Automated identification is subject to the certain inherent limitations so long as the dichotomy to be achieved is essentially random and the number of features involved in the identification is less than twice the number of potential objects to be identified. Such constraints lead to important automated identification system design considerations; the machine capacity must be such that a dichotomy is a practical and feasible goal in the intended environment. Similar results from catastrophe theory show that not all features can be treated as though they had the same metric properties. In addition, the orthogonal feature selection is proposed as an optimization of the feature space with a specified distance measure.

Introduction

The problem of designing a practical automated identification system is essentially one of determining the optimum conditions for and the limitations involved in pattern recognition. In all pattern recognition tasks there is a universe of objects which is to be separated into two or more classes according to testable features. No distinction can be made between physical and virtual objects--that is, between objects which are specified by real-time signal inputs to the identification system and objects which are specified by a particular set of data in some arbitrary storage device within the system.

While there are numerous mathematical results which might be applicable to pattern recognition, it is not always clear how such results should be considered when automated identification systems are being designed. A few results of prime import arise from considerations within catastrophe theory and are concerned with the problems of decidability, clustering, decision rules, distance measures, and parameter selection. While it is not always possible to determine an absolutely correct system design procedure from mathematical results, it is possible to recognize an incorrect design and to know the circumstances under which the system must ultimately fail. Thus, one is led to two possible types of design procedures: one, to design automated identification systems which are not known to be inherently flawed, or two, to design automated identification systems which are flawed while specifying circumstances of well-behaved operation.

Catastrophe Theory and Its Relation to Pattern Recognition

Catastrophe Theory is essentially a set of mathematical methods for predicting the behavior of systems which undergo rapid and (apparently) discontinuous changes of state.* Such systems need not be physical ones. Indeed, formal decidability is a question of the metamathematical behavior of the mathematical system. The unique and exciting result that arises from catastrophe theory is the fact that catastrophes can be classified. While at first it was thought that there were only seven elementary catastrophes in a space of four dimensions,[1] more recent results have shown that, when the difference between the metric properties of space and those of time are taken into account, there are in fact twelve elementary catastrophes.[2] The so-called behavior surfaces of a particular catastrophe tell one a good deal about the system involved: specifically, how to cause, predict, and avoid catastrophic behavior. The mathematical basis of catastrophe theory is far too involved to detail here; however, several volumes are currently available which suffice to explain the methods entailed and these are listed in the references.

The problems of clustering and separability in pattern recognition seem to fall into the domain of catastrophe theory in a rather straightforward manner. In general, there is a universe U of objects which are to be classified into two or more classes C. It is the task of pattern recognition to determine the class membership function f and calculate the class membership of the pattern objects.

* referred to as catastrophes.

We can ask, under the constraints of a given task of classification, whether or not a specific classification is probable or even possible. If the variables of the task (such as the number of objects and the number of adjustable parameters) are changed, the probability of success or failure will change accordingly. When the task dynamics are such that a change from high probability to low probability occurs--generally over a relatively short interval--the system is said to exhibit the behavior of a separability catastrophe.

For example, consider the work of Nilsson[4] concerning the capacity of Φ machines. (A Φ-machine is one which uses Φ functions, these being functions Φ which depend linearly on parameters w_i.) The number of possible dichotomies in a universe of N objects is 2^N. Upon selecting one of these dichotomies at random (with probability 2^{-N}), the probability that it can be implemented by a Φ machine with M + 1 adjustable parameters is obtained by dividing the number of Φ dichotomies by 2^N. Thus we obtain the relations:

$$P_{N,M} = 2^{1-N}\sum_{i=0}^{M}\binom{N-1}{i} \quad N \geq M$$
$$P_{N,M} = 1 \quad \text{for } N < M \tag{1}$$

The probability of implementation, which goes from near impossible to near certainty when the number of objects approximately equals twice the number of adjustable parameters, exhibits the behavior of a separability catastrophe.[3] Separability catastrophes (generally classified as fold catastrophes) are common to the two basic recognition tasks, identification and discrimination. The specific shape of the behavior surface and the nature and number of the degeneracies are dependent only upon the nature of the problem involved--namely, the question of the probability of achieving a selected dichotomy with a system using a given number of adjustable parameters. The quantities involved, however, may be quite different in that the separability catastrophe may occur for a different number of objects relative to the number of adjustable parameters. The determination of the particular behavior surface is straightforward: determine the number of dichotomies which satisfy the constraints of the system—usually linearity--in terms of the number of objects and the number of parameters involved, divide by the total number of dichotomies thus yielding the probability that a randomly chosen dichotomy can be satisfied by the system, and plot the surface which is the relationship between the probability P, the number of parameters M+1, and the ratio λ of the number of objects to the number of parameters.

Unless the number of system allowed dichotomies d is $k2^N$, where k is some constant less than or equal to one, the probability P is dependent upon the value of λ and separability is then less than certain for some value of λ. If k ≤ 1 then P = k ≤ 1. For most automated identification systems, the separability function is a Φ-function and hence the system is a Φ machine. This results in a limit to the usefulness of automated identification system--namely, unless features are chosen in such a way as to force clustering in an n-dimensional feature space, the system is likely to fail as soon as the number of objects to be separated reaches approximately twice the number of adjustable parameters.

If, for example, an automatic voice identification system is designed for use in a population of 200 million people, >100 million weighted parameters are necessary as long as the system is a Φ-machine and therefore uses linear functionals. In systems made to deal with criminological evidence, such a constraint is quite real. If an automated identification system is not large enough--in the sense of a sufficient number of adjustable parameters--then the question of reasonable doubt concerning the identification can be raised legitimately. In the identification task, each dichotomy must separate a single object from all the possible objects. If the region separated from the universe of pattern objects should contain more than one object, then no distinction is possible between the objects contained within that identified region, and the identification fails to be unique. A unique identification is guaranteed if and only if the parameters are chosen such that there is no overlap in the images of the pattern classes in feature space, or the operating environment is limited to a universe of objects with cardinality less than twice the number of adjustable parameters.

Decision Rules

It is advantageous then, based on the preceding results, to choose the features or adjustable parameters in such a way as to reduce the degree to which the images of the pattern classes overlap in the feature space. Even if the Φ machine is capable of changing its parameters, it must have some inherent basis for making the decision and this decision making process is again a pattern recognition task that is subject to the constraints given by Nilsson. The process of choosing appropriate parameters, whether as a set of initial conditions in an iterative adjustment or as a part of the system design, is based on decision rules.

Although no general decision rules are given by catastrophe theory, characteristics are suggested. The selection of a single parameter, to be extracted from either deterministic or probabilistic measures and based on the images of the class patterns, is essentially the choice of a metric or distance measure function on the feature space. The simplest such measure is the Euclidian distance between two samples.

$$d = \sum_{i} \left[(x_i - y_i)^2 \right]^{\frac{1}{2}} \qquad (2)$$

Unfortunately it contains an inherent flaw in that it makes no distinction between spatial and temporal features. Such a distinction, while not a necessary condition in the choice of an appropriate decision rule (and hence appropriate parameters), is important in the evaluation of a system's performance. Should a system fail, in the sense that the probability of separation of the pattern classes is quite low, one would like to know the behavioral characteristics leading to such a failure so that the environmental conditions for well-behaved system performance are then known. As mentioned earlier, catastrophe theory tells us how to determine these conditions for a specific system. It also guarantees that there are no more than seven catastrophe classes in a Euclidean geometry and no more than twelve in a Minkowskian geometry. Because the metric properties of time-like quantities are in fact different from those of space-like quantities, the use of a Euclidean distance is inappropriate to a formal analysis of the behavior of a system. In order to avoid encountering degeneracies in analyzing the catastrophic behavior of a system, one must consistently use a Minkowskian metric in the chosen decision rule. For most purposes, this may be achieved by treating features as complex numbers: a space like (real) quantity, and a time like (imaginary) quantity. For example, the duration of a given parameter is obviously time-like and hence its squared magnitude is a negative quantity in the metric, while amplitude is a space-like quantity with a positive squared value. In general, the resulting metric is given by

$$d^2 = \sum_{i=1}^{k} (\Delta e_i)^2 - \sum_{j=1}^{N+1-k} (\Delta t_j)^2 \qquad (3)$$

where d is the distance measure between two samples, k is the number of space-like parameters, N+1-k is the number of time-like parameters, and Δe and Δt are the differences between corresponding parameters.

In addition to the use of Minkowskian metric, it is desirable to treat the chosen parameters as being independent. The number of adjustable parameters N+1, as previously used, referred only to <u>independent</u> parameters. The probability of achieving a specific dichotomy is near certainty for M less than 2(N+1) only if the parameters are independent in the feature space. This should not be assumed and two tests for independence are useful. Chosen parameters which are understood well enough to be instrumented can be reduced to linear functionals of electrical parameters. These functionals can then be tested in the usual manner to determine if they are indeed independent:

$$\int_{-\infty}^{+\infty} \Phi_1^* \Phi_2 \cdot ds = 0; \qquad (4)$$

where Φ_1 and Φ_2 may be either real or complex functions and if real, then the complex conjugate $\Phi_1^* = \Phi_1$.

If, on the other hand, a mapping from feature-space functionals to electrical parameters is either impractical or even impossible, the independence of the parameters may be tested empirically. This is done by making observations of the parameters and obtaining the commutation relations. If the product of the chosen parameters is order dependent* when measured empirically, then the parameters do not commute and the chosen feature-space is <u>not</u> orthogonal.

<u>Automated Voice Identification System Design</u>

In order to elucidate the applicability of the preceding suggestions, we now turn to a specific example of automated identification system design: voice identification. Two approaches to speaker recognition by machine are more common than others, and we will concentrate on these: the first entails the examination of specific cue material in the form of amplitude-frequency-time matrices and the second entails the extraction of speaker-dependent parameters from the speech signal which are subjected to statistical analysis. In the first approach, the number of weighted parameters of the system may be regarded as the product of the number of matrix cells (i.e., number of frequency bands multiplied by the number of time intervals) and the number of levels used to describe the amplitude. We may assume for the purposes of this discussion that problems of comparison (such as temporal and spatial alligment) are solvable.

The matrix approach is capable of generating relatively large numbers of parameters. Unfortunately the cost of the system is dominated by the cost of memory or information storage. In addition, the amount of time involved in a bit-by-bit comparison of two such matrices can be inordinate if (1) the number of parameters is large, (2) the linear functionals require calculation, and (3) as is generally the case, the decision rule requires further calculation. It is because of these practical aspects that the use of the matrix approach is to be discouraged, especially where near-certain identification is desired in a large speaker population. If one can limit or even estimate the number of potential speakers in the population to which the system will ever be subjected, then a sufficient number of parameters can be chosen. Alternatively, if the number of independent parameters is known, the system user population should be clearly specified by the designer.

* I.e., if for observed parameters A and B, [AB-BA] ≠ 0.

The second approach to automated voice identification, statistical analysis of speaker-dependent parameters extracted from the speech signal, is subject to all the problems with which we have been concerned: separability catastrophe, the choice of independent parameters, and the use of a distance measure appropriate to the chosen parameters. The most critical factor in the statistical approach is the choice of the adjustable parameters. In most cases, this choice is random and the Nilsson result then applies once again--predicting system failure for sufficiently large speaker populations. If the choice of features involves finding invariants, again the problem of an appropriate choice of metric arises. We have determined the signature of the metric and have specified that the parameters be orthogonal (independent) in the space. Unfortunately the weighting of the parameters must still be determined.

Conclusions

In an examination of twenty of the most successful speaker recognition systems as of 1971, none were tested in the region of interest, i.e., where a separability catastrophe should have occurred. The use of Minkowskian geometries in any automated identification system are unknown to this author, and orthogonality conditions are rarely tested. Automated identification systems are unlikely to be generally reliable, either now or in the future, unless specific parameters which yield a separable feature space are discovered.

It may well be that appropriate distance measures or orthogonal parameters will lead to the design of successful and highly reliable automated identification systems over some selected range of populations. For criminological purposes, systems should be designed with limited populations in mind and should never be used in cases where positive identification is a requirement unless it can be shown that populations are sufficiently limited to allow well-behaved system operation.

Acknowledgements

The author would like to express gratitude to Nils Nilsson, of SRI, for his fine publication, Learning Machines; Richard Becker, of SRI, for helpful suggestions and directions; and to Hans Bremermann, of UC-Berkeley, for enlightening discussions of clustering problems.

References

1. R. Thom, "Structural Stability and Morphogenesis," W. A. Benjamin, Inc., Reading, Mass., 1976.

2. G. Wassermann, "(r,s)-Stable Unfoldings and Catastrophe Theory," Structural Stability, the Theory of Catastrophes, and Applications in the Sciences, Springer-Verlag, Heidelberg, Germany, pp. 253-262, 1976.

3. H. Bremermann, "Pattern Recognition by Deformable Prototypes," Structural Stability, the Theory of Catastrophes, and Applications in the Sciences, Springer-Verlag, Heidelberg, Germany, pp. 15-57, 1976.

4. N. J. Nilsson, "Learning Machines," McGraw-Hill Book Company, New York, N.Y., 1965.

5. M. Hecker, et al., "Speaker Recognition," ASHA Monographs, No. 16, 1971.

6. T. Poston and I. Stewart, "Catastrophe Theory and Its Applications," Fearon-Pitman Publishers Inc., Belmont, Calif., 1978.

7. A. Dold and H. Eckmann, ., "Dynamical Systems--Warwick 1974," Springer-Verlag, Heidelberg, Germany, 1975.

8. S. Watanabe, "Frontiers of Pattern Recognition," Academic Press, New York, N.Y., 1972.