Adaptive clustering algorithm

by Lawrence V. O'Malley

Output data from many types of sensor systems (radar, radar warning, sonar, electro-optical, etc.) must be associated with one or more possible sources based on multiple observations of the data. This paper presents an algorithm that associates data with their source by simultaneous *n*-dimensional clustering of multiple data observations. The algorithm first orders the observations by successive nearest neighbor, in the *n*-dimensional Euclidean sense, from a defined starting point. Clusters are then isolated using a method derived from statistical decision theory. The algorithm's primary feature is its ability to perform clustering adaptively without any assumptions about the size, number, or statistical characteristics of the clusters. Since the algorithm was developed for radar warning system processing, a performance comparison with a well-known algorithm used in that field is included.

Introduction

Clustering algorithms attempt to determine "natural" groupings of data and were originally developed for use in biological taxonomy. For example, the birth and death rates for countries of the world [1] can be processed using clustering algorithms to determine whether groups of countries have similar rates. Such clusters may or may not exist. Clustering techniques in general determine whether clusters do indeed exist, and they are applicable to any type of data.

A more specific problem is the detection and isolation, in the presence of measurement noise or uncertainty, of clusters that are known to exist. This problem is encountered in some signal processing applications. For example, a radar warning system on board an aircraft must isolate and identify individual radars by observing a time-interleaved stream of pulsed emissions from many distinct radar sources.

For each received pulse, the system measures and digitizes a set of signal parameter values, such as radio frequency, bearing angle relative to the aircraft, pulse width, and pulse amplitude. The resulting sets of parameter values must be separated in real time into mutually exclusive clusters such that each cluster represents a unique radar source. These clusters are then used to locate and identify the radars.

Figure 1 shows what an ensemble of sources might produce if many received pulses were plotted based on radio frequency and bearing angle. Note that some of these clusters overlap in bearing angle and others in radio frequency. However, they are all distinct if both radio frequency and bearing angle are considered simultaneously. Therefore, a multidimensional clustering algorithm is required to separate them.

The existence of these clusters of parameter values of the radar pulses is known *a priori*. Pulse parameter values from the same source tend to cluster together over the short time interval within which sorting is performed (generally less than ten milliseconds). All radar warning systems now in use and those projected for the future that we know about are based on this assumption. Therefore, in the remainder of this paper we assume the existence of these clusters and are concerned only with the techniques for identifying them.

Radio frequency is measured using various well-known techniques, such as channelized, superheterodyne, or acousto-optical receivers. Bearing angle is measured by using monopulse amplitude comparison or phase interferometry techniques. The cluster spreading is generally modeled using

normal or uniform distributions but cannot be restricted to such idealizations. The spreading is due to measurement errors in the receiving equipment, to variations in the transmitters, and to physical environmental effects (e.g., reflections). The clusters of parameter values of radar pulses can in general be characterized as having a constant unknown mean and constant unknown standard deviation over the time interval of interest.

The solution to the radar warning problem requires a clustering technique in which nothing is assumed about the number and size of clusters to be found. The clusters of parameter values of radar pulses encountered in reality vary from those having a large intracluster spreading of data points and wide separation between clusters to those having little intracluster spread and small separation. The number of clusters is always unknown.

Review of existing clustering techniques

Several well-known algorithms were examined to determine their applicability to the cluster separation problem. The Forgy-Jancey algorithm [2–4] and the K-means algorithm [1] do not meet the requirements of the cluster separation problem because they separate as many clusters as the user desires, whether or not the clusters are meaningful. For real time cluster separation, it is not possible to peruse the data in advance and preset the number of clusters. More important, the problem demands some measure of face validity in that it is desirable to limit the number of clusters isolated to those corresponding to actual signals.

The ISODATA method [5], the minimal spanning tree algorithm [6], and the Leader algorithm [1] do not meet our requirements because of the restrictions on the size of clusters separated.

A new clustering algorithm

The classical clustering techniques do not provide an algorithm flexible enough to meet the requirements of the radar warning problem. The attributes required in a clustering technique for this problem suggested an approach based on statistical decision theory [7, 8].

Each data point in Fig. 1 represents an observation made at a random time, and the set of observations S_1 is seen by a human observer to form a collection of clusters A through F:

$$S_1 = (A \cup B \cup C \cup D \cup E \cup F).$$

Given the set of observations S_1, our algorithm uses a two-phase process to isolate the clusters. The first phase reorders the set S_1; the second phase isolates the clusters in the reordered set using a statistical decision criterion.

- *Phase One*

To reorder set S_1 in the first phase, we select from S_1 the observation that is nearest in Euclidean distance to the origin. This observation (we call it O_1) becomes the first

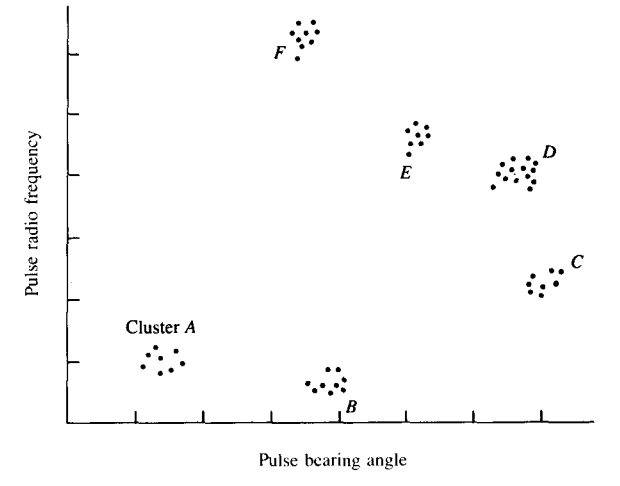

Figure 1

Pulse radar clusters.

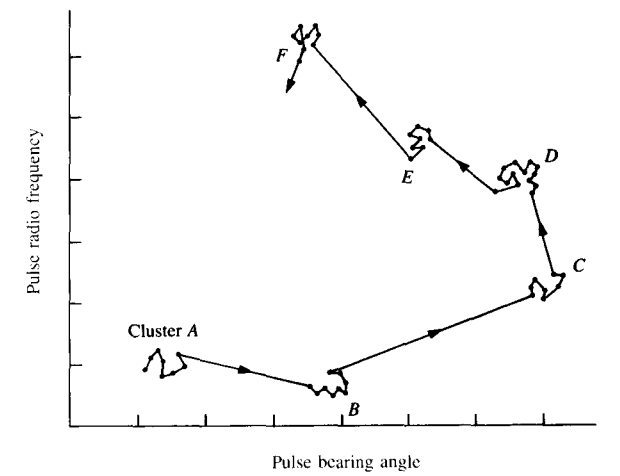

Figure 2

Ordered clusters.

member of the reordered set, which we call S_2. Observation O_1 is then removed from set S_1. Next, the observation in S_1 that is closest in Euclidean distance to O_1, call it O_2, is made the second observation in S_2. It is also removed from S_1. This process is continued until all observations are included in set S_2 and deleted from set S_1. **Figure 2** shows graphically the effect of the reordering done in Phase One.

- *Phase Two*

In the second phase, the set S_2 is processed to isolate clusters A through F in Fig. 2. Since we do not yet know which

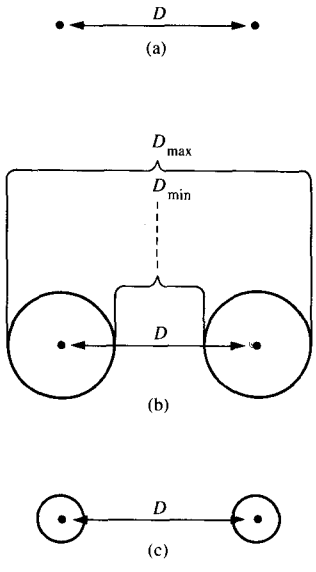

Figure 3

Adjacent clusters with (a) no noise or parameter variation, (b) considerable noise and parameter variation, and (c) moderate noise and parameter variation.

observations belong to each cluster, we define S_2 as it appears to the processor:

$$S_2 = (O_1, O_2, \cdots, O_n),$$

where n is the total number of observations.

Since we do not know the size of the clusters (i.e., the number of observations in each cluster), a statistical decision criterion is used to decide which observations have a high probability of forming a group. To make this decision, we first assume that reliable identification of a cluster requires some minimum number of observations, P. We take four to be the value of P in our application. For other applications, the user of the algorithm must determine a reasonable value for P.

We now form a new temporary set, S_3, comprised of the first $2P$ observations in the ordered set S_2:

$$S_3 = (O_1, O_2, \cdots, O_x),$$

where $x = 2P$.

The Euclidean distance separating each successive pair of observations in the ordered set S_3 is then determined. That pair of adjacent observations whose separation is greatest, say, O_y and O_{y+1}, in the set

$$S_3 = (O_1, O_2, \cdots, O_y, O_{y+1}, \cdots, O_x)$$

may indicate a cluster separation point. The distance between O_y and O_{y+1} is called A_{max}.

In order to determine whether we have indeed found a cluster separation point, we compute the average distance A_{ave} between all other adjacent pairs of observations in S_3, omitting the distance A_{max}:

$$A_{ave} = (D_{O_1,O_2} + D_{O_2,O_3} + \cdots + D_{O_{y-1},O_y} + \cdots + D_{O_{y+1},O_{y+2}} + \cdots + D_{O_{x-1},O_x})/(x - 2).$$

We then compare the ratio A_{max}/A_{ave} to a threshold T. (The method for determining T is described later.) If the ratio is greater than T, we have isolated the cluster of observations O_1 through O_y, which are removed from set S_2, and the temporary set S_3 is discarded. A new set S_3 is then formed from the first $2P$ observations remaining in set S_2, and the process is repeated until the observations in set S_2 have been exhausted.

If a cluster separation point had not been found, observations would be added incrementally to the existing set S_3 from S_2, and the test for a separation point would be repeated.

The ratio test is based on the following rationale. Referring again to Fig. 1, it is evident that the clusters are easily discernible by the human eye, because the separation of one cluster from neighboring clusters is in some sense more significant than the separation of values within a cluster. Phase One provides us with observations ordered such that we can consider one pair of observations at a time and determine whether the distance between them is significant. The distance between clusters can be thought of as a "signal" or the primary indicator of cluster separateness. Consider the case of two adjacent clusters, as in **Figure 3(a)**, where there is no noise or parameter variation. Now assume that considerable noise and parameter value variation are present, as in **Figure 3(b)**. The distance D between clusters is obscured and could be estimated to be between D_{min} and D_{max}.

In **Figure 3(c)**, we see the same situation except that the spreading of observations is considerably less. It is easier to decide that we have two separate clusters in Fig. 3(a) than in Fig. 3(c), and in Fig. 3(c) than in Fig. 3(b), because in each case we have less noise. Thus, the problem of separating two clusters can be treated as a classical detection problem, i.e., a statistical decision can be used [7, 8]. The ratio A_{max}/A_{ave} is thus a ratio of "signal" to "noise," signal being the maximum intercluster distance and noise the average intracluster distance.

We use a Neyman-Pearson criterion to determine the significance threshold T, because we assume that we do not know anything about the *a priori* probabilities of the cluster distributions. The Neyman-Pearson criterion provides a means for setting the threshold of the ratio test.

This is done by using the conditional probabilities P_f (probability of false alarm) and P_d (probability of detection). Here P_f is the probability of deciding that two clusters are

separate when they are not (i.e., an error), and P_d is the probability of deciding that two clusters are separate when they are. The Neyman-Pearson criterion constrains P_f to be less than a selected value and maximizes P_d under this constraint.

The procedure of increasing the set S_3 by one new observation restricts the algorithm to one cluster pair separation at a time, because otherwise intercluster distances from other cluster pairs would perturb the ratio test on a given cluster pair.

Outliers (measurements having very large errors) can occur due to a transient failure in a measurement mechanism or any number of other causes. By their very definition, they are low probability occurrences and as such have only temporary effect. If data points caused by this phenomenon occur midway between clusters, they can cause clusters not to separate. Techniques exist to preprocess the data to eliminate outliers, image processing techniques using FFTs, for example. The decision to deal directly with outliers or to depend upon their low probability of occurrence is dependent upon a given application.

All clustering algorithms have in common the question of scaling or normalization when the coordinate axes consist of parameters whose numerical values have widely different magnitudes (e.g., radio frequency in thousands of megahertz and bearing angle in degrees from 0 to 360). The scaling is done to prevent the larger axis from dominating the distance calculation.

The observations are normalized before ordering in Phase One. For Phase Two, we employ a scaling technique which normalizes data by the maximum measured values within a cluster pair. The normalization is restricted to the cluster pairs defined by the set S_3. The scaling is done each time a new observation is added to the subset. We use this technique because it prevents measurements from other clusters from distorting the scaling of a given cluster pair.

Clustering using parameters which can be measured in a modulo sense (e.g., bearing angle is measured modulo 360) presents a special case. It is obvious that clusters whose measurements straddle the 0 to 360 crossover require special attention to account for the anomalies. Axis shifting or replication of data by adding 360 are two well-known techniques for dealing with this situation.

Performance analysis

The algorithm was simulated to determine its performance relative to the optimal theoretical performance and to that of an existing algorithm. The existing algorithm is a comparison tolerance algorithm widely used in sorting radar warning system pulse data. It is similar to the Leader algorithm [1].

The algorithms were first tested against a hypothetical set of two clusters to determine how well each algorithm separated the clusters. The clusters were modeled as one-dimensional normal distributions whose means are separated by a distance S. The standard deviations of each distribution were equal. For a trial, four samples were drawn from each distribution and each algorithm performed a sort of the samples. If all samples were correctly associated with the distribution from which they originated, then the trial was reported as a success. The probability of sorting correctly was computed for one thousand trials. The distance S between the means was varied from one to six standard deviations, and the trials were repeated. The adaptive algorithm had its threshold value set using the Neyman-Pearson criterion.

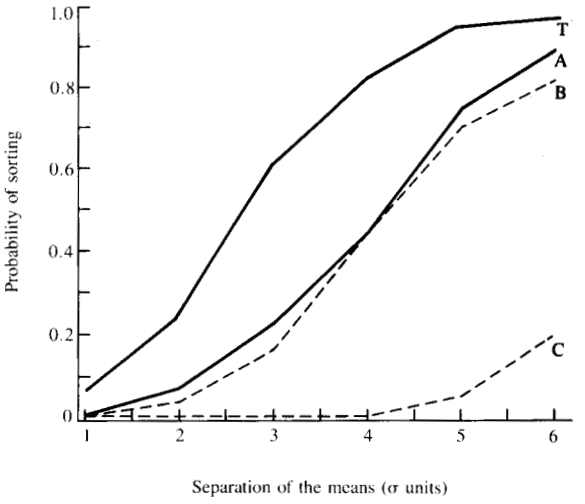

Figure 4

One-dimensional performance.

The comparison tolerance algorithm was run for two limiting cases. It had a tolerance selected on the basis of the expected standard deviation of all distributions. This tolerance was usually set to three times the standard deviation for the largest distribution expected. If the distributions to be sorted exactly matched the tolerance, the performance of the comparison tolerance algorithm was considered to be at its maximum. If the distributions to be sorted had a smaller or larger standard deviation than the tolerance, the performance was considered to be poorer. The limiting cases for the comparison tolerance algorithm are the best case (i.e., a tolerance of three standard deviations) and the worst case (i.e., a tolerance of six standard deviations). This bounds the performance of the comparison tolerance algorithm.

The optimal theoretical sorting of the two distributions could be performed with a classical hypothesis test if the statistical characteristics of the distributions and their mean separation were known. This performance is plotted in **Figure 4** as Curve T. Curve A represents the performance of

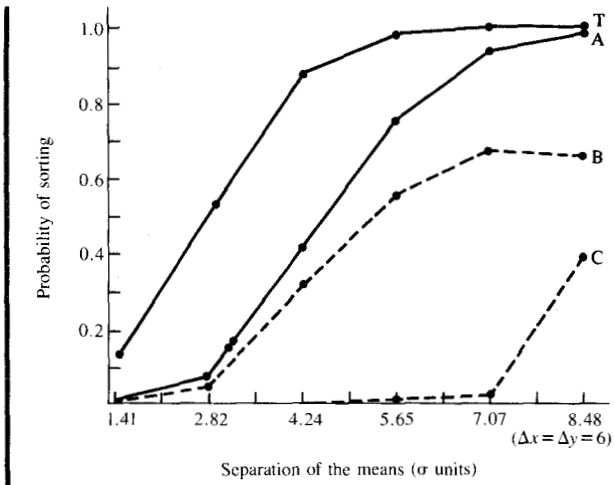

Figure 5

Two-dimensional performance.

the adaptive algorithm. Curves B and C represent the range of performance of the comparison tolerance algorithm from best to worst case, respectively. It is clear that the performance of the adaptive clustering algorithm is better than that of the comparison tolerance algorithm even in its optimal case (i.e., Curve B).

The two-dimensional case was also simulated, and the results are shown in **Figure 5**. Curves T, A, B, and C represent the same performance measures as in Fig. 4. The performance of the adaptive clustering algorithm is much better than for the one-dimensional case, and the advantage over the comparison tolerance algorithm is more pronounced.

Conclusions

A method for n-dimensional clustering of radar warning system pulse data has been presented. Although examples can be postulated in which this algorithm fails to order observations properly or fails to cluster properly, its performance has been shown to be superior to the most widely used current technique. The primary advantage of the algorithm is its ability to adapt to radar source signatures. Many of today's systems have signal processing algorithms which use tolerances. These tolerances are functions of assumptions about a particular operational environment and the peculiarities of the system components. Since the adaptive algorithm is tolerance independent, future signal processing designs using such an algorithm can be made more universal.

Further work is needed on the algorithm in the following areas:

1. Development of a theoretical model to more extensively characterize its performance. Particularly important is the study of larger sample sizes and higher dimensions.
2. Comparison of the algorithm with other clustering techniques.
3. Simulation of a real world radar pulse clustering problem to prove feasibility.

Acknowledgments

This work had its genesis in a question proposed to me by R. F. Osbahr. R. E. Poupard provided enthusiasm, encouragement, and technical assistance in exploring all the subtleties of the algorithm. J. Jephson suggested the technique for incremental application of the algorithm. R. E. Blahut, W. Vanderkulk, and T. Cochrane provided assistance in many discussions. T. Hooks performed much of the simulation programming.

References

1. J. A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, Inc., New York, 1975.
2. R. C. Jancey, "Multidimensional Group Analysis," *Aust. J. Bot.* **14**, No. 1, 127–130 (April 1966).
3. E. W. Forgy, "Cluster Analysis of Multivariate Data: Efficiency vs Interpretability of Classifications," *Biometrics Abst.* **21**, No. 3, 768–769 (September 1965).
4. B. Black, M. Arozullah, and W. Ladew, "Modeling of Shadows in Radar Clutter," *Defense Technical Information Center Technical Report No. ADA 089702*, Defense Logistics Agency, Cameron Station, Alexandria, VA 22314, July 1980.
5. G. H. Ball and D. J. Hall, "ISODATA, A Novel Method of Data Analysis and Pattern Classification," Technical Report, Stanford Research Institute (now SRI International), Menlo Park, CA, May 1965.
6. C. T. Zahn, "Graph Theoretical Methods for Detecting and Describing Gestalt Clusters," *IEEE Trans. Computers* **C-20**, 68–86 (January 1971).
7. M. I. Skolnik, *Introduction to Radar Systems*, McGraw-Hill Book Co., Inc., New York, 1962.
8. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley & Sons, Inc., New York, 1968.

Received September 8, 1983; revised July 27, 1984

Lawrence V. O'Malley *IBM Federal Systems Division, Route 17C, Owego, New York 13823.* Mr. O'Malley is an advisory engineer in the Electronic Defense Systems Engineering Department. He is presently involved in the systems analysis and design of various defense programs. He joined IBM in 1968 at the Owego laboratory and has worked on many electronic warfare programs, including the S-3A Antisubmarine Warfare System, the AN/APR-38 Wild Weasel Radar Defense Suppression System, and the Passive Identification and Detection Set for the U.S. Navy Tomahawk Cruise Missile. Mr. O'Malley is presently involved in the development of signal processing architectures and algorithms. He received a B.S. in mathematics in 1968 from the University of Scranton, Pennsylvania, and an M.S. in computer systems in 1975 from the State University of New York at Binghamton. Mr. O'Malley is a member of the Association of Old Crows and the Institute of Electrical and Electronics Engineers.