Cromley, R. and S. Morse. [1988]. "Some Comments on Numerical Line Generalization," Geographical Analysis, 20, 263-69.

Douglas, D. and T. Peucker. [1973]. "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature," The Canadian Cartographer, 10, 112-22.

Marino, J. [1979]. "Identification of Charactersitic Points Naturally Occurring Lines: An Empirical Study," The Canadian Cartographer, 16, 70-80.

McMaster, R. [1983]. "A Mathematical Evaluation of Simplification Algorithms," Proceedings, AUTO CARTO 6, 2, 267-76.

McMaster, R. [1986]. "A Statistical Analysis of Mathematical Measures for Linear Simplification," The American Cartographer, 13, 103-16.

McMaster, R. [1987a]. "Automated Line Generalization," Cartographica, 24, 74-111.

McMaster, R. [1987b]. "The Geometric Properties of Numerical Generalization," Geographical Analysis, 19, 330-46.

Peucker, T. [1976]. "A Theory of the Cartographic Line," International Yearbook of Cartography, 16, 134-43.

Robinson, A., R. Sale, J. Morrison, and P. Muehrcke. [1984]. Elements of Cartography, Fifth Edition, New York: John Wiley.

Toregas, C., C. ReVelle, R. Swain, and L. Bergman, "The Location of Emergency Service Facilities," Operations Research, 19, 1366-73.

White, E. [1985]. "Assessment of Line-Generalization Algorithms Using Characteristic Points," The American Cartographer, 12, 17-27.

A COMPARISON OF STATISTICAL AND ADAPTIVE PATTERN RECOGNITION APPROACHES TO LINE FEATURE IDENTIFICATION

James E. Mower
Department of Geography and Regional Planning
147 Social Sciences
State University of New York at Albany
Albany, New York 12222 USA
jem97@ALBNY1VX

Abstract

In this paper, the author will compare the relative benefits and problems introduced by the application of competitive learning and Bayesian statistical approaches to line feature clustering. The paper will examine the cognitive aspects of pattern recognition (comparing issues of similarity in feature space to similarity in descriptor space) as well as computational issues associated with the implementation of Bayesian and competitive learning procedures on sequential and parallel processing environments.

Introduction

The usefulness and beauty of a map largely depends upon the quality of its linework. Quality is reflected in the appropriate selection, placement, and execution of line features. Current laser printing and plotting devices render examples of automated linework that are often indistinguishable from those of a trained cartographer.

The automatic selection and placement of features is more problematic and remains an area of active research. Effective representations of line features are distinguished by their identification of critical points along the line (Marino 1979).

Most current line simplification procedures require an operator to choose a tolerance value that guides the selection of critical points. Buttenfield (1987) suggests that line feature categories be established with associated tolerance values. The selection of a tolerance value then becomes a pattern classification problem.

This study compares the application of competitive learning procedures and standard cluster analysis procedures to the automatic clustering of cartographic lines. Competitive learning is one of several basic parallel distributed processing (PDP) mechanisms that have been applied to pattern recognition problems (Rumelhart and others 1986). Samples of digitized linework, represented by sets of Fourier descriptors, are submitted to an ART2 network (Carpenter and Grossberg 1987) and the SAS CLUSTER procedure. Cluster assignments are compared. Issues of computational efficiency and cognition are discussed as they apply to the clustering procedures.

Clustering and Classification

Classification procedures assign observations to pre-existing categories. Clustering procedures take a set of observations and group them on measures of multivariate similarity, without regard to *a priori* class definitions. This study will use clustering procedures to suggest future line taxonomies and to serve as a preliminary line feature/attribute lookup table.

Describing the Shape of Cartographic Lines

Representing Lines

Cartographic lines can be described in raster and vector formats. Several studies extract information from chain encoded raster formats (Freeman 1974). Thapa (1988) detects critical points in chain encoded lines by convolving them with their masks and searching for zero crossings. Kuhl and Giardina (1982) use chain encoded lines to extract their elliptical Fourier descriptors. Freeman (1977) measures lengths and areas of "peninsulas" and "bays" to determine critical points along chain encoded lines.

Many more studies develop shape descriptors from vector line representations. Pavlidis (1977) and Brady (1983) review numerous techniques for the description of open and closed curves. Techniques for converting between raster and vector representations are well-known and can be found in Foley and Van Dam (1982) and Zahn (1966).

Methods for Describing Lines

Line description methods vary in their input requirements. Some, like the symmetric axis transform (SAT) and methods based upon the Fourier transform, operate upon closed lines. Others, like Freeman's method for the detection of critical points (1977), operate upon open curves.

This study will use the Fourier descriptors of closed line segments as developed in Mower (1990). Fourier descriptors have the following advantages:

(1) they are invariant under graphic transformations;

(2) they provide a compact description of shape; and

(3) they are easy to compute.

Generating Fourier Descriptors of Cartographic Lines

Zahn and Roskies (1972) discuss a method for extracting sets of Fourier descriptors for planar closed curves. A closed curve is represented by vertices V_0 through V_m. Starting at V_0, lengths of arcs between successive vertices are accumulated, as are the angles between successive arcs. The sum of the angles will be 2Π radians unless the curve spirals, in which case the sum will be greater than 2Π radians. Zahn and Roskies suggest equations (1), (2), and (3) for generating the Fourier coefficients a_n and b_n for closed curves. We can use these coefficients as shape descriptors which, in polar coordinates, represent the nth harmonic amplitude and phase angle respectively.

$$a_n = -1 / (n / \pi) \sum_{k=1}^{m} (\Delta\Phi_k \sin [(2\pi n l_k) / L] \quad (1)$$

$$b_n = 1 / (n / \pi) \sum_{k=1}^{m} (\Delta\Phi_k \cos [(2\pi n l_k) / L] \quad (2)$$

where

$$l_k = \sum_{i=1}^{k} \Delta l_i \quad (3)$$

In formulas (1), (2), and (3), m is the number of vertices on the closed curve, $\Delta\Phi_k$ is the change in angle of the line at vertex k, L is the total length of the line, and Δl_i is the length of the line segment between vertices l and l+1.

Approaches to Line Pattern Clustering

Cluster Analysis

Rohlf and Archie (1984) establish a mosquito wing shape taxonomy derived from a cluster analysis of digitized wing samples. Using Fourier coefficients as shape descriptors, the authors digitized photographs of the right wings of 127 species of mosquitoes. They computed a matrix of Fourier coefficients for the first 15 harmonics. This matrix was submitted to several multivariate analysis procedures, including cluster analysis.

The authors found that cluster members were visually similar but that the overall clustering pattern was not closely related to preexisting taxonomies. They also state that individual Fourier coefficients are unlikely to be morphologically meaningful.

For the first 15 harmonics, the authors found the shape contribution of the odd harmonics to be much lower than that of the even harmonics. In their study of dual axis Fourier shape analysis for cartographic forms, Moellering and Rayner (1984) found that the coefficients of the first 5 harmonics explained 99% of the variance in the shape of the island of Hokkaido. Zahn and Roskies (1972), operating on hand-printed character sets, found that the coefficients from 10 harmonics were sufficient to reconstruct visually recognizable numerals. They cite work by Brill (1967) who found that approximately seven harmonics were sufficient for discriminating between five numerals.

Competitive learning techniques

Pao (1989) and Rumelhart and others (1987) review developments in parallel distributed processing (PDP) that have led to the creation of procedures for learned clustering and classification of novel input patterns.

PDP studies the application of massively parallel processing networks to information processing. Individual processors in a network compute simple functions of their input, broadcasting their output values through weighted links to other processors as inhibitory or excitatory signals. A set of weights typically maps multiple input patterns to unique output patterns.

A recurring theme in PDP processing is the generation of appropriate response patterns to presentations of stimulus patterns. These pattern associators "learn" to generate response patterns through repeated exposure to stimulus-response pairs. Similarly, auto associators learn to identify a pattern from a degraded input sample.

This study will examine PDP regularity detectors, or competitive learning networks, that cluster novel input patterns without regard to *a priori* categories. Competitive learning networks are composed of an input layer and an output layer. If a processor in the input layer detects a stimulus in the input pattern, it broadcasts an activation value through weighted links to each of the processors in the output layer. The output processor receiving the greatest summed activation values wins the competition with the other output processors, redistributing a higher proportion of its total weights along the activated input lines. Processors that do not win competitions do not redistribute their weights. If the input patterns are structured, processors in the output layer will consistently win competitions for "similar" patterns. In this sense, competitive learning networks cluster their input patterns.

Carpenter and Grossberg (1987) discuss the development of adaptive resonance theory (ART) and its application to clustering problems. ART clustering procedures add the concept of a vigilance parameter to clustering, preventing clusters from claiming input patterns if the measured distance between the pattern and the cluster center is greater than the vigilance parameter value. The authors have developed ART procedures to cluster binary (ART1) and analog (ART2) input patterns.

Cluster Analysis and PDP

Everitt (1974) generalizes clustering as the grouping of similar entities and the separation of dissimilar entities with respect to the scores of the entities on a set of variables. Clusters may be formed hierarchically, with larger clusters subsuming smaller ones or they may represent a partitioning of the entities into mutually exclusive regions.

Everitt notes that definitions for cluster and similarity are often vague. As a result, it is difficult to identify the number of clusters present in a data set.

During each iteration, a hierarchical clustering procedure assigns entities and clusters to larger clusters. The investigator is responsible for determining the appropriate number of clusters. Competitive learning networks partition the pattern space, "assigning" novel patterns to processors in the output layer.

McClelland and Rumelhart (1988) note that competitive learning procedures may not produce stable cluster assignments if input patterns lack structure. ART networks control pattern oscillation through the vigilance parameter. As the value of the parameter is increased, the number of clusters will increase and cluster size will decrease.

Competitive learning networks are inherently parallel constructions. Carpenter and Grossberg (1987) note that the performance of an ART network does not degrade significantly as the number of patterns increases, assuming that the number of available processors in the output layer remains large relative to the number of input patterns.

Most cluster analysis procedures operate on sequential computing architectures. As such procedures are rewritten for parallel architectures, their expected running time will decrease considerably. It is necessary, then, to compare cluster analysis and competitive learning procedures on the quality of their output rather than on the speed of its production.

Cartographic Line Clustering

Recent studies have proposed techniques for clustering linework. Buttenfield (1987) identifies 5 variables for describing cartographic lines and uses these

descriptors to classify line samples. Mower (1988) develops parallel algorithms for implementing Buttenfield's classification procedure in matrix and neural network programming models. Mower (1990) uses Fourier descriptors and ART2 procedures to cluster digitized linework.

Method

This study uses the set of line samples digitized for Mower (1990), taken from 14 USGS 1:250,000 topographic maps. The set consists of 39 sections of varied coastlines and 5 sections of rivers. 11 of the coastlines were sections of the Lake Ontario and Lake Michigan shorelines, 3 were barrier beaches along the south shore of Long Island, New York, 9 were sections of the mid and northern Maine coastlines, 4 were selected from New York state Finger Lakes shorelines, 6 were sections of the Chesapeake Bay coastline, 3 were portions of the South Carolina and Georgia coastlines, 1 was a section of the Washington coast, and 2 were sections of the Virginia Atlantic coastline.

The 5 rivers included 1 section each of the Colorado River in the Grand Canyon, the Genesee River in New York, the lower Hudson River, and the Pearl and Black Rivers in Mississippi and Louisiana.

Fourier coefficients cannot be extracted from open curves. Zahn and Roskies (1972) solve this problem for stroked characters by first performing an outline trace of the component strokes and then extracting coefficients from the trace.

This study uses a polygonization procedure to convert the rivers and coastlines to closed curves. The Fourier coefficients for the first 10 harmonics are extracted from the polygonized linework using equations (1), (2), and (3).

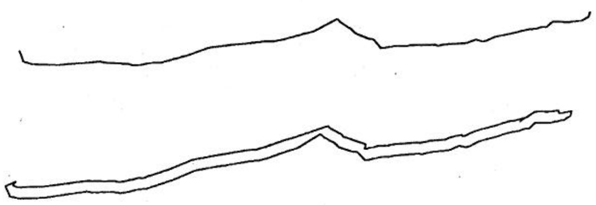

Figure 1. A line sample (top) and its polygonized form

ART2 Procedure

Mower (1990) submitted the raw Fourier coefficients to an ART2 simulator, provided by Arun Rao. For a detailed description of the simulator, see Rao and others (1989) and Mower (1990). The simulator assigns each line sample to a cluster, represented by a logical processor. An input layer reads pattern information and outputs weighted values to the cluster processors (the output layer). A top-down vigilance parameter prevents the assignment of patterns to clusters when the cluster activation falls below the vigilance threshold. If no cluster is similar enough to the sample (that is, if no cluster is activated above the vigilance threshold), a new cluster is formed.

Clustering Procedure

Everitt (1974) notes that strongly correlated variables can promote spurious clustering. Correlations of the Fourier coefficients showed that the A_n and the A_{n+1}

coefficients were strongly correlated. To examine the effect of these correlations on the clustering procedure, a principal components analysis was run on the Fourier coefficients. None of the coefficients loaded highly on any of the first six components. A cluster analysis based upon the first six components differed markedly from an analysis conducted on the raw coefficients, producing visually dissimilar clusters. No further analysis was conducted on the principal components.

The raw coefficients were submitted to the SAS CLUSTER procedure using the average linkage method and Ward's minimum variance method. Both methods perform hierarchical clustering. Cluster agglomeration was halted at 10 on inspection of pair-wise RMS distance values across joined clusters.

Results

ART2 Cluster Assignments

Mower (1990) found that the Rao ART2 simulator produced several, moderately sized clusters with relatively similar appearing lines at a vigilance parameters of .006. When the vigilance parameter was strengthened, the most visually dissimilar samples consistently dropped out of larger clusters. However, the stronger values also assigned a large proportion of the total samples to unique clusters.

Although several of the ART2 clusters were visually homogeneous, the geomorphological processes associated with the lines were often quite different. It was found that the ART2 cluster assignments were not geomorphologically based.

SAS Cluster Assignments

Ward's minimum variance method tends to produce similar sized clusters (SAS Institute 1985). For the 10 line sample clusters, cluster 1 and cluster 2 assignments dominate the remaining clusters (16 and 9 assignments respectively). However, each of the remaining clusters, except for cluster 10, hold 2 or 3 samples. Cluster 10 holds 1 sample. The average linkage method does not produce such regular cluster sizes. Cluster 1 assignments dominate the other clusters (27 assignments). The remaining cluster assignments fall to 4 for cluster 2, 3 for clusters 3 and 5, 2 for cluster 4 and 1 for clusters 6 through 10.

Ward's method produces clusters that are more visually homogeneous than those produced by the average linkage method. Although all but 1 of the samples that Ward's method assigns to cluster 1 are also clustered together by the average linkage method, many of the 2 and 3 member clusters differ in their groupings.

The average linkage method groups samples together that are sometimes visually dissimilar. This is most apparent in cluster 1. Not surprisingly, all but one of the samples from the Atlantic beaches (from Long Island, New York to South Carolina) and all of the New York State Finger Lakes shorelines are assigned to cluster 1. However, 2 samples from the Maine coast and 1 from the Chesapeake Bay shoreline are also assigned to cluster 1. In cluster 4, a section of the Maine coastline is grouped with a section of the Lake Michigan shoreline. Although the Lake Michigan sample is not smooth, it is much less crenulated than the sample from Jonesport, Maine.

Average linkage does produce several clusters of lines that are reasonably similar in appearance. Cluster 2 contains 2 samples from the Chesapeake Bay shoreline, a section from the Genesee river and a section of Maine coastline. Although visually similar, no clear geomorphological process dominates the cluster. Most of the other clusters lack common processes as well.

Ward's method produces a more visually homogeneous clustering of Atlantic beaches and Finger Lakes shorelines (cluster 1). Unlike cluster 1 in the average linkage method, only about half of the Great Lakes shorelines are assigned to cluster 1; the other half are assigned primarily to cluster 2. Cluster 2 is relatively visually homogeneous, except for the inclusion of a section from the Camden, Maine coastline. Of the 9 sections of Maine coastline, this section is probably the least crenulated.

Ward's method produces many small clusters with good visual similarities. These examples are displayed in Figures 2a through 2d.

Comparison of ART2 and SAS Cluster Assignments

The ART2 and SAS procedures generate relatively good visual clusters of the line samples. However, none of the clustering procedures produce clusters that are geomorphologically consistent.

Rohlf and Archie (1984) found that clusters of mosquito wing shapes based upon Fourier coefficients ignored structural characteristics. It is likely that Fourier coefficients are generally not well-suited to the extraction of process-based clusters.

If cluster assignments are not grounded in geomorphology, then evaluations of the clustering methods must be based on other measures of shape or cognitive measures of similarity. Moellering and Rayner (1982) suggest several other measures of shape for cartographic analysis.

Conclusion and Suggestions for Further Research

This study has compared competitive learning and cluster analysis methods for the clustering of cartographic line samples. We have found that both methods produce reasonable visual clusters of line samples, yet do not incorporate knowledge of geomorphological process.

A number of extensions to this study may be examined before abandoning the use of Fourier coefficients as process oriented shape descriptors. The number of samples can be increased to include more representatives of well-defined processes. As such information is added, the study will move from a clustering problem to a classification problem. *A priori* knowledge of class distributions will allow us to develop and evaluate new shape discriminants.

We have not shown a clear difference between the results of a cluster analysis procedure or that of an ART2 procedure. For line clustering, the main advantage of an ART2 procedure is its insensitivity to variations in the number of known clusters. The author intends to extend this study to the evaluation of other shape descriptors for cartographic line clustering and classification and to explore matrix parallel algorithms for identifying linear features.

Figure 2a. Ward's method. Both samples are from Chesapeake Bay

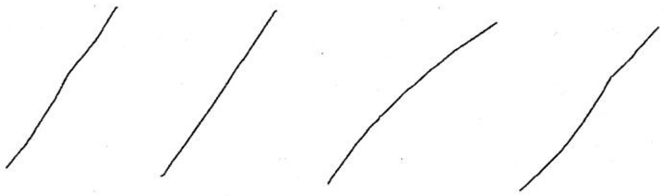

Figure 2b. Ward's method. The 1st, 2nd, and 4th samples are from the Atlantic Coast. The 3rd sample is from a Great Lakes shoreline

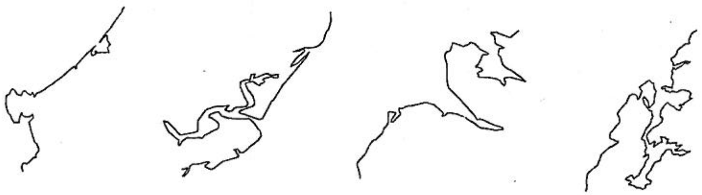

Figure 2c. Ward's method. From left to right, a Great Lakes shoreline, an Atlantic coastline, a section of the north shore of Long Island, and a section of the Maine coast

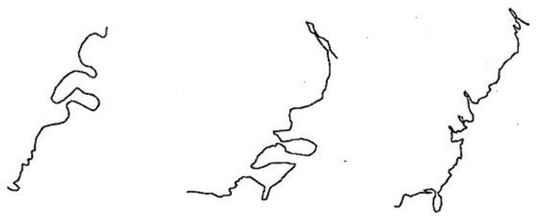

Figure 2d. Ward's method. From the left, a river and two sections of the Maine coast

References

Brady, M. 1983. "Criteria for the Representations of Shape." Human and Machine Vision, ed. J. Beck and A. Rosenfeld, pp. 39-84. New York: Academic Press.

Buttenfield, B.P. 1987. "Automating the Identification of Cartographic Lines." The American Cartographer 14, no. 1:7-20.

Carpenter, G.A. and S. Grossberg 1987. "ART 2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns." Applied Optics 26, no. 23:4919-4930.

Everitt, B. 1974. Cluster Analysis. New York: John Wiley and Sons.

Foley J.D. and A. Van Dam 1982. Fundamentals of Interactive Computer Graphics. Reading, MA: Addison-Wesley Publishing Company.

Freeman, H. 1974. "Computer Processing of Line-Drawing Images." ACM Computing Surveys 6, no. 1:57-97.

Freeman, H. 1977. "Shape Description via the Use of Critical Points." Pattern Recognition 10:159-166.

Kuhl, F.P. and C.R. Giardina 1982. "Elliptical Fourier Features of a Closed Contour." Computer Graphics and Image Processing 18:236-258

Marino, J.S. 1979. "Identification of Characteristic Points along Naturally Occurring Lines/an Empirical Study." The Canadian Cartographer 16, no.1:70-80.

McClelland, J.L. and D.E. Rumelhart 1988. Explorations in Parallel Distributed Processing. Cambridge, MA: The MIT Press.

Moellering, H. and J.N. Rayner 1982. "The Dual Axis Fourier Shape Analysis of Closed Cartographic Forms." The Cartographic Journal 19, no. 1:53-59.

Mower, J.E. 1988. "A Neural Network Approach to Feature Recognition along Cartographic Lines." GIS/LIS'88 Proceedings. Falls Church, VA: American Society for Photogrammetry/American Congress on Surveying and Mapping.

Mower, J.E. 1990. "Line Feature Clustering through Competitive Learning." Unpublished manuscript.

Pavlidis, T. 1977. Structural Pattern Recognition. Berlin: Springer-Verlag.

Rao, A., M.R. Walker, L.T. Clark, and L.A. Akers 1989. "Integrated Circuit Emulation of ART1 Networks." Proceedings, First IEE International Conference on Artificial Neural Networks.

Rohlf, J.F. and J.W. Archie 1984. "A Comparison of Fourier Methods for the Description of Wing Shape in Mosquitoes (Diptera: Culicidae)." Systematic Zoology 33, no. 3:302-317.

Rumelhart, D.E., J.L. McClelland, and the PDP Research Group 1986. Parallel Distributed Processing, Volume 1: Foundations. Cambridge, MA: MIT Press.

SAS Institute 1985. SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: SAS Institute, Inc.

Thapa, K. 1988. "Critical Points Detection and Automatic Line Generalization in Raster Data Using Zero-Crossings." Cartographic Journal 25, no.1:58-68.

Zahn, C.T. 1966. "Two Dimensional Pattern Description and Recognition via Curvaturepoints." Stanford Linear Accelerator Center, Stanford University, Stanford, CA, Technical Report SLAC-70.

Zahn, C.T. and R.Z. Roskies 1972. "Fourier Descriptors for Plane Closed Curves." IEEE Transactions on Computers 21, no. 3:269-281.

A PARALLEL IMPLEMENTATION OF FRANKLIN'S UNIFORM GRID TECHNIQUE FOR LINE INTERSECTION DETECTION ON A LARGE TRANSPUTER ARRAY

Sara Hopkins and Richard G. Healey
Regional Research Laboratory for Scotland
Department of Geography, University of Edinburgh
Drummond Street, Edinburgh, Scotland
e-mail: sara@uk.ac.ed.geovax, rgh@uk.ac.ed.geovax

Abstract

The increased availability and popularity of parallel computers is of considerable importance for the GIS community because of the ever increasing demands for intensive processing of large data sets. This paper investigates the design and implementation of the first stage of a parallel polygon overlay algorithm, namely line intersection detection. The algorithm employed, which uses Franklin's uniform grid technique is discussed, together with issues specific to its implementation on a transputer array, an MIMD distributed memory parallel computer.

Introduction

Polygon overlay is central to many kinds of GIS analysis, but it is also frequently the source of major bottlenecks in processing, particularly when large coverages, stored in vector form, are to be overlaid. The computationally intensive nature of overlay arises from two sources. The first of these is the fundamental combinatorial complexity of the operation (Saalfeld 1989), with a requirement to compare many line segments with many others for the purpose of locating intersections, even though this task can be reduced in magnitude by strategies to localize the comparisons made (e.g. White 1978). The second is that overlay is a multi-stage operation, involving not only line intersection detection but also linkage of intersected arcs from the overlay map coverages to form new polygons and the assignment of attributes from the overlay coverage polygons to these new polygons.

The combined requirement for computationally intensive localized and multi-stage processing suggests that vector polygon overlay operations could be enormously accelerated by implementation on parallel processing machines using a combination of geometric decomposition and algorithmic parallelization techniques (Bowler et al. 1987). This paper describes the design and implementation of an algorithm for parallel line intersection detection on a large parallel machine, using a uniform grid technique. The work is part of a larger investigation into parallel polygon overlay methods. An algorithm for the polygon linkage stage has also been developed, but this will be the subject of a future paper.