MAPPING MULTIVARIATE SPATIAL RELATIONSHIPS FROM REGRESSION TREES BY RECURSIVE BINARY PARTITIONS OF THE SPECTRUM

Denis White
Department of Geosciences
Oregon State University
Corvallis, OR 97331
USA
Tel: 503-754-4476
Fax: 503-737-1200
Email: denis@heart.cor.epa.gov

Multivariate thematic mapping is a special challenge for cartographic design. Many national atlases, for example, include complex maps with multiple variables or themes. Computer mapping and geographic information systems make construction of multivariate maps quite easy. Are there good cartographic design methods for such maps?

Displaying the results of multivariate statistical analyses is one type of multivariate mapping. Classification and regression trees (CART) are one type of multivariate statistical analysis. In a CART, the observation space, or the collection of places for which data are recorded, is successively partitioned into a prediction tree. At each node in the tree the CART algorithm searches for the value of one of the independent variables that explains the greatest amount of variation in the dependent variable. The observations are split into two groups at each node according to this splitting criterion. Although the tree could grow until each observation has its own "leaf", in practice the tree is "pruned" to a size that contains the maximum gain in predictive power. Each leaf of the pruned tree then consists of a unique set of observations. CART is especially useful, and generally more robust than multiple linear regression, for example, when the multivariate relationships are nonlinear, and when missing values may be a significant factor in analysis.

What is an appropriate and effective cartographic method to display the spatial relationships revealed by a regression tree? First, we observe that each leaf in the tree has a specific and unique prediction relationship; that is, each leaf has its own set of independent variables, with corresponding value ranges, that predicts the value of the dependent variable at the observations belonging to the leaf. Second, we can arrange the tree such that observations with lower values of the splitting independent variable are always on the left at each node. This provides an unambiguous ordering to the tree.

Since each leaf has an unambiguous ordered position in the tree, yet a unique composition of predictor variable splits, one method for assigning mapping symbols to the observations of the leaf is by locating the leaf in a corresponding position along the color spectrum. To locate positions on the spectrum we successively bisect it as the prediction tree grows. In the final pruned tree then, observations that are closer in spectral hue to others indicate a closer relationship in the structure of the prediction variables. I illustrate this regression tree mapping method with examples from biogeography and other applications.

ICC '95 Topic: Digital thematic cartography