A Test of an Unsupervised Machine Learning Procedure Applied to Cloud Classification Data

Diana Gordon
Naval Research Laboratory
Washington, DC 20375-5337
and
Paul M. Tag and Richard L. Bankert
Naval Research Laboratory
Monterey, CA 93943-5502

ABSTRACT

Machine learning algorithms can be subdivided into two types-supervised and unsupervised. Usually, supervised learning is the more practical and useful technique; using various features as input, data samples have known outcomes that the user wants to predict. However, unsupervised learning is appropriate when the user does not know the subdivisions into which the data samples, using relevant predictor features, should be divided. Prior categorical division may not be obvious because the problem may be a new one, for which the user has little experience. In such a case, an unsupervised learning procedure can provide insight into groupings that may make physical sense and facilitate future analysis.

To explore the potential of an unsupervised learning procedure, we have applied a program called Autoclass to a meteorological data set of samples for which there are known outcomes. Autoclass was developed by Cheeseman, Kelly, Self, Stutz, Taylor, and Freeman at the NASA Ames Research Center and is based upon Bayesian classification principles. Given a database of attribute vectors and a class model, Autoclass finds a maximum posterior probability classification.

Our data set was developed from satellite imagery of cloud regions that were expertly labeled into ten classes. Because cloud types hold meteorological significance, an automated classification from satellite imagery is of obvious use. The data set we are using is composed of 1633 labeled cloud samples, each of which has 204 calculated features (textural, spectral, and physical). Supervised learning methods have produced theoretical classification accuracies of over 85% for this data set. The purpose of this experiment is to compare cloud classes produced from an unsupervised learning procedure to traditional cloud classes. The differences may provide insight into alternate ways of viewing or interpreting clouds.