CHAPTER

27

CLASSIFICATION TECHNIQUES FOR REMOTELY SENSED DATA

Eldho Varghese and Grinson George

Fishery Resources Assessment Division
ICAR-Central Marine Fisheries Research Institute

1. Introduction

Hyperspectral imaging sensors measure the radiance of the materials within each pixel area at a very large number of contiguous spectral wavelength bands. So, they can generate hundreds of images of a scene on the real surface. The radiance is converted into hyperspectral data cube digital form. The spectral information available in a hyperspectral image (cube) may serve to classify the nature of the target object because every material had a unique fixed spectrum and could be used as a spectral signature of the material and perhaps provide additional information for further processing and exploitation. Hyperspectral data contain extremely rich spectral attributes, which offer the potential to discriminate more detailed classes with classification accuracy.

2. Cluster Analysis (Unsupervised learning)

Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses. The most common unsupervised learning method is cluster analysis, which is used for exploratory data analysis to find hidden patterns or grouping in data.

Cluster analysis is usually done in an attempt to combine cases into groups when the group membership is not known prior to the analysis. Cluster analysis is a technique for grouping individual or objects into unknown groups.

There are 4 basic steps to conduct cluster analysis for any data. Those are given below:

1. Select a suitable distance measure.
2. Select a clustering algorithm.
3. Determine the number of clusters.
4. Validate the analysis.

2.1 Clustering Methods (Johnson and Wichern, 2006)

The commonly used methods of clustering fall into two general categories.

(i) Hierarchical and

(ii) Non hierarchical.

2.1.1 Hierarchical cluster Analysis

Hierarchical clustering techniques proceed by either a series of mergers or a series of successive divisions. Agglomerative hierarchical method starts with the individual objects, thus there are as many clusters as objects. The most similar objects are first grouped and these initial groups are merged according to their similarities. Eventually, as the similarity decreases, all subgroups are fused into a single cluster.

Divisive hierarchical methods work in the opposite direction. An initial single group of objects is divided into two sub groups such that the objects in one sub group are far from the objects in the others. These subgroups are then further divided into dissimilar subgroups. The process continues until there are as many subgroups as objects *i.e.*, until each object form a group. The results of both agglomerative and divisive method may be displayed in the form of a two dimensional diagram known as Dendrogram. It can be seen that the Dendrogram illustrate the mergers or divisions that have been made at successive levels.

Linkage methods are suitable for clustering items, as well as variables. This is not true for all hierarchical agglomerative procedure. The following types of linkage are now discussed:

(i) Single linkage (minimum distance or nearest neighbour),

(ii) Complete linkage (maximum distance or farthest neighbour) and

(iii) Average linkage (average distances).

Also other methods of hierarchical clustering techniques like Ward's method and Centroid method are available in the literature.

Steps of Agglomeration in Hierarchical Cluster analysis

The following are the steps in the agglomerative hierarchical clustering algorithm for groups of N objects (items or variables).

i. Start with N clusters, each containing a single entity and an N×N symmetric matrix of distance (or similarities) **D** = $\{d_{ik}\}$.

ii. Search the distance matrix for the nearest (most similar) pair of clusters. Let the distance between most similar clusters U and V be $d_{uv.}$

iii. Merge clusters U and V. Label the newly formed cluster (UV). Update the entries in the distance matrix by (a) deleting the rows and columns corresponding to clusters U and V and (b) adding a row and column giving the distances between cluster (UV) and the remaining clusters.

iv. Repeat steps (ii) and (iii) a total of N-1 times (All objects will be in a single cluster after the algorithm terminates). Record the identity of clusters that are merged and the levels (distances or similarities) at which the mergers take place.

2.1.2 Non Hierarchical Clustering Method

Non-hierarchical clustering techniques are designed to group items, rather than variables, into a collection of K clusters. The number of clusters, K, may either be specified in advance or determined as part of the clustering procedure. Because a matrix of distance does not have to be determined and the basic data do not have to be stored during the computer run. Non hierarchical methods can be applied to much larger data sets than can hierarchical techniques. Non hierarchical methods start from either (1) an initial partition of items into groups or (2) an initial set of seed points which will form nuclei of the cluster.

2.1.2.1 K means Clustering (Afifi, Clark and Marg, 2004)

The K means clustering is a popular non hierarchical clustering technique. For a specified number of clusters K the basic algorithm proceeds in the following steps:

- Divide the data into K initial cluster. The number of these clusters may be specified by the user or may be selected by the program according to an arbitrary procedure.
- Calculate the means or centroid of the K clusters.
- For a given case, calculate its distance to each centroid. If the case is closest to the centroid of its own cluster, leave it in that cluster; otherwise, reassign it to the cluster whose centroid is closest to it.
- Repeat step (iii) for each case.
- Repeat steps (ii), (iii), and (iv) until no cases are reassigned.

2.2. Dendrogram

Dendrogram is also called hierarchical tree diagram or plot, and shows the relative size of the proximity coefficients at which cases are combined. The bigger the distance coefficient or the smaller the similarity coefficient, the more clustering involved combining unlike entities, which may be undesirable. Cases showing low distance are close, with a line linking them a short distance from the left of the Dendrogram, indicating that they are agglomerated into a cluster at a low distance coefficient, indicating alikeness. When, on the other hand, the linking line is to the right of the Dendrogram the linkage occurs at a high distance coefficient, indicating the cases/clusters were agglomerated even though much less alike.

2.3. Distance Measures

Some distance measures commonly used for assessing spectral similarity/dissimilarity are as follows:

1) Spectral Similarity Index or Spectral Correlation or Spectral Angle Mapper (SAM)
2) Spectral Contrast Angle
3) Spectral Information Divergence (SID)
4) Spectral Absorption Index (SAI)
5) Euclidian Distance
6) Mahalanobis D^2
7) City-Block Distance

2.4. Validations

- **Relative clustering validation**, which evaluates the clustering structure by varying different parameter values for the same algorithm (e.g., varying the number of clusters k). It's generally used for determining the optimal number of clusters.
- **External clustering validation**, which consists in comparing the results of a cluster analysis to an externally known result, such as externally provided class labels. Since we know the "true" cluster number in advance, this approach is mainly used for selecting the right clustering algorithm for a specific dataset.
- **Internal clustering validation**, which use the internal information of the clustering process to evaluate the goodness of a clustering structure without reference to external information. It can be also used for estimating the number of clusters and the appropriate clustering algorithm without any external data.
- **Clustering stability validation**, which is a special version of internal validation. It evaluates the consistency of a clustering result by comparing it with the clusters obtained after each column is removed, one at a time.

3. Dicriminant Function Analysis (Supervised learning)

Discriminant function analysis is a statistical analysis to predict a categorical dependent variable (called a grouping variable) by one or more continuous or binary independent variables (called predictor variables). The original dichotomous discriminant analysis was developed by Sir Ronald Fisher in 1936. Discriminant function analysis is useful in determining whether a set of variables is effective in predicting category membership. Discriminant analysis is used when groups are known a priori (unlike in cluster analysis).

Each case must have a score on one or more quantitative predictor measures, and a score on a group measure. In simple terms, discriminant function analysis is classification - the act of distributing things into groups, classes or categories of the same type.

The assumptions of discriminant analysis are the same as those for MANOVA. The analysis is quite sensitive to outliers and the size of the smallest group must be larger than the number of predictor variables. The major assumptions are:

- Multivariate normality: Independent variables are normal for each level of the grouping variable.
- Homogeneity of variance/covariance (homoscedasticity): Variances among group variables are the same across levels of predictors. Can be tested with Box's M statistic.
- It has been suggested, however, that linear discriminant analysis be used when covariances are equal, and that quadratic discriminant analysis may be used when covariances are not equal.
- Multicollinearity: Predictive power can decrease with an increased correlation between predictor variables.
- Independence: Participants are assumed to be randomly sampled, and a participant's score on one variable is assumed to be independent of scores on that variable for all other participants.
- It has been suggested that discriminant analysis is relatively robust to slight violations of these assumptions, and it has also been shown that discriminant analysis may still be reliable when using dichotomous variables (where multivariate normality is often violated).

Discriminant analysis works by creating one or more linear combinations of predictors, creating a new variable for each function. These functions are called discriminant functions. The number of functions possible is either N_g-1 where N_g = number of groups, or p (the number of predictors), whichever is smaller. The first function created maximizes the differences between groups on that function. The second function maximizes differences on that function, but also must not be correlated with the previous function. This continues with subsequent functions with the requirement that the new function not be correlated with any of the previous functions.

Summary

Unsupervised learning is a heuristic technique for classifying cases into groups when knowledge of the actual group membership is unknown. Unless there is considerable separation among the inherent group, it is not realistic to expect very clear results with

unsupervised learning. In particular if the observations are distributed in a nonlinear manner, it may be difficult to achieve distinct groups. Cluster analysis is quite sensitive to outliers. The data should be carefully screened before running cluster programs.

Discriminant analysis is a multivariate technique concerned with classifying distinct set of objects (or set of observations) and with allocating new objects or observations to the previously defined groups. It involves deriving variates, which are combination of two or more independent variables that will discriminate best between a priori defined groups.

References

Afifi, A., Clark, V. A. and Marg, S. (2004). *Computer Aided Multivariate Analysis*. USA, Chapman & Hall.

Chatfield, C. and Collins, A. J. (1990). *Introduction to Multivariate Analysis*. Chapman and Hall Publications.

Hair, J. F., Anderson R. E., Tatham, R. L. and Black, W. C. (2006). *Multivariate Data Analysis.* 5th Edn., Pearson Education Inc.

Johnson, R. A. and Wichern, D. W. (2006). *Applied Multivariate Statistical Analysis.* 5th Edn., London, Inc. Pearson Prentice Hall.

Timm, N. H. (2002). *Applied Multivariate Analysis*. 2nd Edn. , New York, Springer-Verlag

Anderson, T. W., 1984, An Introduction to applied Multivariate Statistical Analysis, *John Wiley & Sons,* New York.