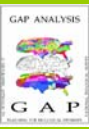




Tools for Applying Classification Trees for Land Cover Mapping



John Lowry, Christine Garrard and Vinod Chowdary
RSGIS Laboratory, College of Natural Resources, Utah State University
Logan, Utah 84322-5275



ABSTRACT:

The use of Classification Trees (CT) for land cover mapping is not new, and in fact is becoming increasingly common. Classification Trees (sometimes called Decision Trees, or CART (Classification And Regression Trees)) offer several advantages over classification algorithms traditionally used for land cover mapping. A major advantage is the ability to effectively use both categorical and continuous predictor datasets with different measurement scales. Other advantages include the ability to handle non-parametric training data, good computational efficiency and an intuitive hierarchical representation of discrimination rules.

The CT algorithm uses multiple explanatory variables to predict a single response variable. In the case of land cover, explanatory variables may include spectral bands and their transformations, DEM derived data, and rasterized vector data, such as geology or soils. The algorithm determines the appropriate characteristics of the response variable by recursively splitting the predictor data into increasingly more homogenous groups (figure 1) with the result of producing a hierarchical tree composed of "rules" (figure 2). The CT algorithm is a *data mining* tool and is available from several commercial vendors including Insightful (S-PLUS™), RuleQuest Research (See5/C5.0©) and Salford Systems (CART®).

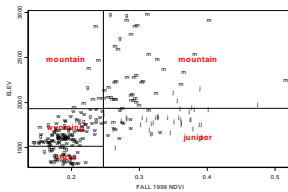


Figure 1. Example of recursive splitting on 2 variables



Figure 2. Example of "decision tree" from figure 1

A major challenge in using Classification Trees for land cover mapping is the spatial application of rules generated from the CT software within a geographic information system. Through the course of the Southwest Regional GAP land cover mapping effort, Utah State University developed and/or evaluated three tools for spatially applying CT rules. This poster briefly presents these tools. *StatMod* is an ArcView™ extension that interfaces with S-PLUS to build classification trees as ArcInfo™ Grids. The *Imagine CART Module* was developed by Earth Satellite Corp. to allow the integration of See5/C5.0 output with ERDAS Imagine. Finally *RuleMaker* is a web-based script allowing a user to translate the tree output from S-PLUS into an Imagine .mdl file.

SELECTED REFERENCES:

Breiman, L., J. H. Friedman, R. A. Olshen and C. J. Stone. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth.

Brown de Colstoun, E. C., M. H. Story, C. Thomposon, K. Commisso, T. G. Smith and J.R. Irons. 2003. National park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sens. Environ.* 85: 316-327.

Friedl, M. A. and C. E. Brodley. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 61: 399-409.

Hansen, M., R. Dubyah and R. DeFries. 1996. Classification trees: An alternative to traditional land cover classifiers. *Int. J. Remote Sensing* 17(5): 1075-1081.

Lawrence, R. L. and A. Wright. 2001. Rule-based classification systems using classification and regression tree (CART) analysis. *Photogrammetric Engineering & Remote Sensing*, 67(10): 1137-1142.

Pal, M. and P. M. Mather. 2003. An assessment of effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* 86: 554-565.

Verbyla, D. L. 1987. Classification trees: a new discrimination tool. *Can. J. For. Res.* 17: 1150-1152.

STATMOD Zone ArcView Extension

Software Requirements: ArcView Spatial Analyst and S-PLUS.

Mode of Use: The extension allows you to submit jobs automatically to S-PLUS. This means the user may set all Classification Tree parameters, such as pruning and cross-validation criteria, within an ArcView dialog box (figure 3). These are submitted to S-PLUS where the tree is created, then output from S-PLUS automatically returns to ArcView for spatial application. The final output is an ArcInfo grid. The response dataset may be either a point or polygon shapefile or coverage with an attribute containing the response value. Explanatory datasets may be shapefiles, ArcInfo grids or coverages.



Figure 3. Classification & Regression Trees dialog box. In this example, the Value fields from the Aspect, Azch, Elevation, and Precipitation themes will be used as independent variables for a classification tree model. The dependent variable will be the Value field from the Samples point theme.

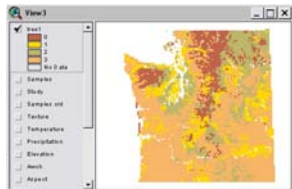


Figure 4. Example of classification tree model results. Four different output classes are predicted.

Other Features: StatMod Zone is capable of generating both classification trees and regression trees using S-PLUS. If the user chooses, there is the ability to edit tree rules before spatially applying them. By default a classification map is produced, however, a user may also produce a probability or "confidence" map. The confidence map is a spatial representation of the probabilities associated with each cell being mapped correctly. The extension also interfaces with the SAS statistical software for spatial application of logistic regression models.

Author and Availability: StatMod Zone was developed by Christine M. Garrard in partial fulfillment of an M.S. in Biology at Utah State University. The ArcView extension is free. The extension and user guide can be downloaded from the following URL: <http://www.gis.usu.edu/~chrisg/avext/>

Imagine CART Module

Software Requirements: ERDAS Imagine and See5/C5.0.

Mode of Use: The module was developed using ERDAS Imagine EML and C-toolkit. Use of the model involves three steps (figure 5). First, the user uses a CART module dialogue box to create the input for the classification tree software. The response dataset is an Imagine .img file with numeric values corresponding to response attributes. The explanatory datasets must be Imagine .img files. This step produces a summary of the data in the proper file format for See5/C5.0. The second step involves using See5/C5.0 to generate the classification tree (i.e. decision tree). Once the See5/C5.0 tree file is produced, the user returns to the CART module in Imagine to spatially apply the classification tree rules generated in See5/C5.0. The output is an Imagine .img file.

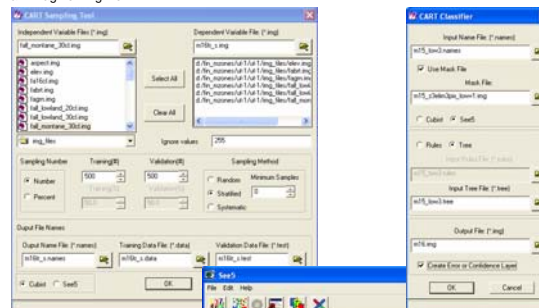


Figure 5. Step one, use CART Module to sample explanatory datasets in preparation for See5. Step two, run See5 to create Classification Tree. Step three, use output from See5 in CART Module to generate Classification Tree.

Other Features: The module can also generate regression trees in a similar fashion. However to create regression trees the software Cubist (also by RuleQuest Research) must be used. Using See5/C5.0 it is possible to use the procedure called "Boosting" which has been shown to improve the results produced from classification trees. In addition to a classification map, the user has the option to produce a confidence map, which is a spatial representation of the probabilities associated with each pixel being mapped correctly.

Author and Availability: The Imagine CART Module was developed by Earth Satellite Corporation for the USGS, Eros Data Center in support of the National Land Cover Database (NLCD) 2000. The Imagine CART Module software can be downloaded for free from: <http://www.gis.usu.edu/%7Eregap/download/C5Module/>

RuleMaker Web-based Perl Script

Software Requirements: ERDAS Imagine and S-PLUS.

Mode of Use: RuleMaker is a web-based perl script that can be used to translate the tree output from S-PLUS into an Imagine .mdl file (figure 6). The user must first use S-PLUS to create the classification tree. The tree file must be saved in the print tree format. The user goes to the website at <http://www.gis.usu.edu/docs/projects/swgap/rulemaker.html>. RuleMaker will require the user to upload the tree output from S-PLUS, a variables file, the path to where you wish to place the .mdl file, and an output image name. Once RuleMaker is run (by pressing submit). The resulting .mdl file is downloaded to your computer and is ready to use in Imagine. Spatially applying the rules generated from RuleMaker is done by running Imagine's Modeler, producing an Imagine .img file.

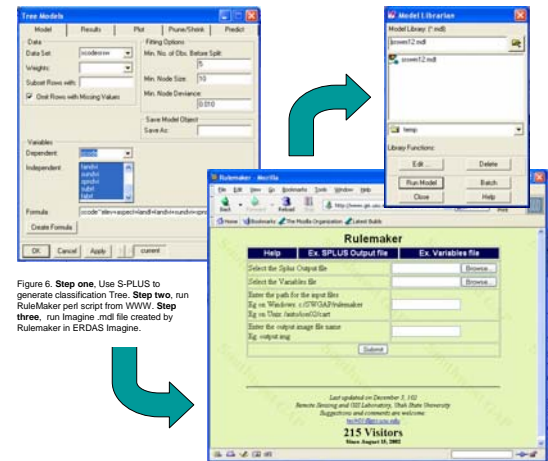


Figure 6. Step one, use S-PLUS to generate classification tree. Step two, run RuleMaker perl script from WWW. Step three, run Imagine .mdl file created by RuleMaker in ERDAS Imagine.

Other Features: RuleMaker provides the option of producing a confidence map, which is a spatial representation of the probabilities associated with each pixel being mapped correctly. Since the Imagine .mdl file is an editable text file, the rules can be edited by the user.

Author and Availability: RuleMaker was developed by Vinod Chowdary in partial fulfillment of an M.S. in Computer Science at Utah State University. You may access RuleMaker on the web for free at:

<http://www.gis.usu.edu/docs/projects/swgap/rulemaker.html>