See also: Invasive Species.

Further Reading

Bellows TS and Fisher TW (eds.) (1999) *Handbook of Biological Control: Principles and Applications of Biological Control*. San Diego: Academic Press.

Clausen CP (ed.) (1978) *Agricultural Research Service: Handbook No. 480: Introduced Parasites and Predators of Arthropod Pests and Weeds: A World Review*. Washington, DC: USDA: Agricultural Research Service.

DeBach P and Rosen D (1991) *Biological Control by Natural Enemies*. Cambridge, UK: Cambridge University Press.

Follett PA and Duan JJ (eds.) (2000) *Nontarget Effects of Biological Control*. Boston, MA: Kluwer Academic.

Jervis M and Kidd N (eds.) (1996) *Insect Natural Enemies: Practical Approaches to their Study and Evaluation*. London: Chapman and Hall.

Julien MH and Griffiths MW (eds.) (1998) *Biological Control of Weeds, a World Catalogue of Agents and their Target Weeds*, 4th edn. Wallingford: CABI Publishing.

Van Driesche J and Van Driesche RG (2000) *Nature Out of Place: Biological Invasions in a Global Age*. Washington, DC: Island Press.

Van Driesche RG, Hoddle M, and Center T (2008) *Control of Pests and Weeds by Natural Enemies, an Introduction to Biological Control*. London: Blackwell.

Classification and Regression Trees

G G Moisen, US Forest Service, Ogden, UT, USA

Introduction

Frequently, ecologists are interested in exploring ecological relationships, describing patterns and processes, or making spatial or temporal predictions. These purposes often can be addressed by modeling the relationship between some outcome or response and a set of features or explanatory variables. Some examples from ecology include:

- analyzing bioclimatic factors affecting the presence of a species in the landscape,
- mapping forest types from remotely sensed data,
- predicting forest attributes over large geographic areas,
- identifying suitable wildlife habitat,
- making sense of complex ecological data sets with hundreds of variables,
- predicting microhabitat affecting fish species distributions,
- developing screening tests for unwanted plant species,
- monitoring and mapping landcover change through time,
- using environmental variables to model the distribution of vegetation alliances,
- assessing biological indicators of environmental factors affecting fish habitat, and
- identifying fuels characteristics for fire spread models.

Modeling ecological data poses many challenges. The response as well as the explanatory variables may be continuous or discrete. The relationships that need to be deciphered are often nonlinear and involve complex interactions between explanatory variables. Missing values for both explanatory and response variables are not uncommon, and outliers almost always exist. In addition, ecological problems usually demand methods that are both easy to implement and easy to interpret. Frequently, many different statistical tools are employed to handle unique problems posed by the various scenarios. This diverse set of tools might include multiple or logistic regression, log linear models, analysis of variance, survival models, and the list continues. Classification and regression trees, however, offer a single tool to work with all these challenges. This article

describes classification and regression trees in general, the major concepts guiding their construction, some of the many issues a modeler may face in their use, and, finally, recent extensions to their methodology. The intent of the article is to simply familiarize the reader with the terminology and general concepts behind this set of tools.

Overview of the Fitting Process

Classification and regression trees are intuitive methods, often described in graphical or biological terms. A tree is typically shown growing upside down, beginning at its root. An observation passes down the tree through a series of splits, or nodes, at which a decision is made as to which direction to proceed based on the value of one of the explanatory variables. Ultimately, a terminal node or leaf is reached and predicted response is given.

Trees partition the explanatory variables into a series of boxes (the leaves) that contain the most homogeneous collection of outcomes possible. Creating splits is analogous to variable selection in regression. Trees are typically fit via binary recursive partitioning. The term binary refers to the fact that the parent node will always be split into exactly two child nodes. The term recursive is used to indicate that each child node will, in turn, become a parent node, unless it is a terminal node. To start with, a single split is made using one explanatory variable. The variable and the location of the split are chosen to minimize the impurity of the node at that point. There are many ways to minimize the impurity of each node. These are known as splitting rules. Each of the two regions that result from the initial split are then split themselves according to the same criteria, and the tree continues to grow until it is no longer possible to create additional splits or the process is stopped by some user--defined criteria. The tree may then be reduced in size using a process known as pruning.

Assigning a predicted value to the terminal nodes can be done in a number of ways. Typically, in classification trees, values at the terminal nodes are assigned the class which represents the plurality of cases in that node. The rules of class assignment can be altered based on a cost function, to adjust for the consequences of making a mistake for certain classes, or to compensate for unequal sampling of classes. In the case of regression trees, values at the terminal node are assigned using the mean of cases in that node.

As an example, consider the problem of modeling the presence or absence of the tree species *Pseudotsuga menziesii* (Douglas fir) in the mountains of northern Utah using only information about elevation (ELEV) and aspect (ASP), where data take the form:

Douglas fir	*Elevation (m)*	*Aspect*
Absent	2045	E
Present	2885	SE
Present	2374	NE
Absent	2975	S
...	...	...

Figure 1 illustrates a simple classification tree for this problem. Beginning with all 1544 observations at the root, the 393 cases that fall below an elevation of 2202 m are classified as having no Douglas fir. If elevation is greater than 2202 m, as is the case for 1151 observations, then more information is needed. The next split is made at an elevation of 2954 m. These very-high-elevation observations above the cutoff are also classified as having no Douglas fir. Turning now to the remaining 928 moderate-elevation observations, yet more fine-tuning is needed. The third split occurs at an elevation of 2444 m. The 622 moderately high elevation cases above 2444 m are classified as having Douglas fir present. The final split uses aspect to determine if Douglas fir is likely to grow on the remaining 306 moderately low sites, predicting Douglas fir to be present on the cooler, wetter northerly and easterly slopes, and absent on the hotter, dryer exposures.

At a minimum, construction of a tree involves making choices about three major issues. The first choice is how splits are to be made: which explanatory variables will be used and where the split will be imposed. These are defined by splitting rules. The second choice involves determining appropriate tree size, generally using a pruning process. The third choice is to determine how application-specific costs should be incorporated. This might involve decisions about assigning varying misclassification costs and or accounting for the cost of model complexity.

Splitting Rules

Binary recursive partitioning, as described above, applies to the fitting of both classification and regression trees. However, the criteria for minimizing node impurity (i.e., maximizing homogeneity) are different for the two methods.

For Regression Trees

For regression trees, two common impurity measures are:

Least squares. This method is similar to minimizing least squares in a linear model. Splits are chosen to minimize the sum of squared error between the observation and the mean in each node.

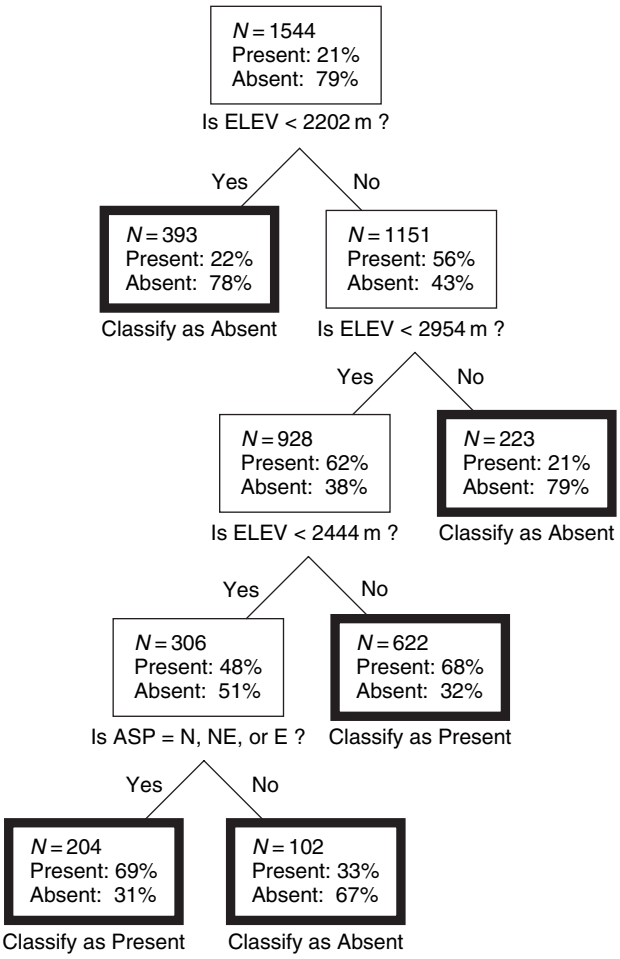

Figure 1 A simple example of a classification tree describing the relationship between presence/absence of *P. menziesii* and explanatory factors of elevation (ELEV) and aspect (ASP) in the mountains of northern Utah. Thin-lined boxes indicate a node from which a split emerges. Thick-lined boxes indicate a terminal node.

Least absolute deviations. This method minimizes the mean absolute deviation from the median within a node. The advantage of this over least squares is that it is not as sensitive to outliers and provides a more robust model. The disadvantage is in insensitivity when dealing with data sets containing a large proportion of zeros.

For Classification Trees

There are many criteria by which node impurity is minimized in a classification problem, but four commonly used metrics include:

Misclassification error. The misclassification error is simply the proportion of observations in the node that are not members of the majority class in that node.

Gini index. Suppose there are a total of K classes, each indexed by k. Let $\hat{p}_{mk}$ be the proportion of class k observations in node m. The Gini index can then be written as $\sum_{k=1}^{K} \hat{p}_{mk}\left(1-\hat{p}_{mk}\right)$. This measure is frequently used in practice, and is more sensitive than the misclassification error to changes in node probability.

Entropy index. Also called the cross-entropy or deviance measure of impurity, the entropy index can be written $\sum_{k=1}^{K} \hat{p}_{mk}\log \hat{p}_{mk}$. This too is more sensitive than misclassification error to changes in node probability.

Twoing. Designed for multiclass problems, this approach favors separation between classes rather than node heterogeneity. Every multiclass split is treated as a binary problem. Splits that keep related classes together are favored. The approach offers the advantage of revealing similarities between classes and can be applied to ordered classes as well.

Pruning

A tree can be grown to be quite large, almost to the point where it fits the training data perfectly, that is, sometimes having just one observation in each leaf. However, this

results in overfitting and poor predictions on independent test sets. A tree may also be constructed that is too small and does not extract all the useful relationships that exist. Appropriate tree size can be determined in a number of ways. One way is to set a threshold for the reduction in impurity measure, below which no split will be made. A preferred approach is to grow an overly large tree until some minimum node size is reached. Then prune the tree back to an optimal size. Optimal size can be determined using an independent test set or cross-validation (described below). In either case, what results is a tree of optimal size accompanied by an independent measure of its error rate.

Independent Test Set

If the sample size is sufficiently large, the data can be divided into two subsets randomly, namely, one for training and other for testing. Defining sufficiently large is problem specific, but one rule of thumb in classification problems is to allow a minimum of 200 observations for a binary classification model, with an additional 100 observations for each additional class. An overly large tree is grown on the training data. Then, using the test set, error rates are calculated for the full tree as well as all smaller subtrees (i.e., trees having fewer terminal nodes than the full tree). Error rates for classification trees are typically the overall misclassification rate, while for regression problems, mean squared error or mean absolute deviation from the median are the criteria used to rank trees of different size. The subtree with the smallest error rate based on the independent test set is then chosen as the optimal tree.

Cross-Validation

If the sample size is not large, it is necessary to retain all the data for training purposes. However, pruning and testing must be done using independent data. A way around the dilemma is through v-fold cross-validation. Here, all the data are used to fit an initial overly large tree. The data is then divided into (usually) $v = 10$ subgroups, and 10 separate models fit. The first model uses subgroups 1–9 for training, and subgroup 10 for testing. The second model uses groups 1–8 and 10 for training, and group 9 for testing, and so on. In all cases, an independent test subgroup is available. These 10 test subgroups are then combined to give independent error rates for the initial overly large tree which was fit using all the data. Pruning of this initial tree proceeds as it did in the case of the independent test set, where error rates are calculated for the full tree as well as all smaller subtrees. The subtree with the smallest error rate based on the independent test set is then chosen as the optimal tree.

Questions often arise as to whether one should use an independent test set or cross-validated estimates of error rates. One thing to consider is that cross-validated error rates are based on models built with only 90% of the data. Consequently, they will not be as good as a model built with all of the data and will consistently result in slightly higher error rates, providing the modeler a conservative independent estimate of error. However, in regression tree applications in particular, this overestimate of error can be substantially higher than the truth, giving more incentive to the modeler to find an independent test set.

1-SE Rule

Under both the testing and cross-validation sections above, tree size was based on the minimum error rate. A slight modification on this strategy is often used where the smallest tree size is selected such that the error rate is within one standard error of the minimum. This results in more parsimonious trees, with little sacrifice in error.

Costs

The notion of costs is interlaced with the issues of splitting criteria and pruning, and is used in a number of ways in fitting and assessing classification trees.

Costs of Explanatory Variables and Misclassification

In many applications, some explanatory variables are much more expensive to collect or process than others. Preference may be given to choosing less expensive explanatory variables in the splitting process by assigning costs or scalings to be applied when considering splits. This way, the improvement made by splitting on a particular variable is downweighted by its cost in determining the final split.

Other times in practice, the consequences are greater for misclassifying one class over another. Therefore, it is possible to give preference for correctly classifying certain classes, or even assigning specific costs to how an observation is misclassified, that is, which wrong class it falls in.

Cost of Tree Complexity

As discussed in the pruning section, an overly large tree can easily be grown to some user-defined minimum node size. Often, though, the final tree selected through tree pruning is substantially smaller than the original overly large tree. In

the case of regression trees, the final tree may be 10 times smaller. This result can be a substantial amount of wasted computing time. Consequently, one can specify a penalty for cost complexity which is equal to the resubstitution error rate (error obtained using just the training data) plus some penalty parameter multiplied by the number of nodes. A very large tree will have a low misclassification rate but high penalty, while a small tree will have a high misclassification but low penalty. Cost complexity can be used to reduce the size of the initial overly large tree grown prior to pruning, which can greatly improve computational efficiency, particularly when cross-validation is being used.

One process that combines the cross-validation and cost complexity ideas is to generate a sequence of trees of increasing size by gradually decreasing the penalty parameter in the cost-complexity approach. Then, tenfold cross-validation is applied to this relatively small set of trees to choose the smallest tree whose error falls within one standard error of the minimum. Because each time a tenfold cross-validation procedure is run a modeler might see a different tree size chosen, multiple (like 50) tenfold processes may be run, with the most frequently appearing tree size chosen.

Additional Tree-Fitting Issues

Although the main issues of fitting classification and regression trees revolve around splitting, pruning, and costs, numerous other details remain. Several of these are discussed below.

Heteroscedasticity

In the case of regression trees, heteroscedasticity, or the tendency for higher-value responses to have more variation, can be problematic. Because regression trees seek to minimize within-node impurity, there will be a tendency to split nodes with high variance. Yet, the observations within that node may, in fact, belong together. The remedy is to apply variance-stabilizing transformations to the response as one would do in a linear regression problem. Although regression trees are invariant to monotonic transformations on explanatory variables, transformations like a natural log or square root may be appropriate for the response variable.

Linear Structure

Classification and regression trees are not particularly useful when it comes to deciphering linear relationships, having no choice but to produce a long line of splits on the same variable. If the modeler suspects strong linear relationships, small trees can first be fit to the data to partition it into a few more similar groups, and then standard parametric models can be run on these groups. Another alternative available in some software packages is creating linear combinations of the explanatory variables, then entering these as new explanatory variables for the tree.

Competitors and Surrogates

It should be noted that when selecting splits, classification and regression trees may track the competitive splits at each decision point along the way. A competitive split is one that results in nearly as pure a node as the chosen split. Classification and regression trees may also keep track of surrogate variables. Use of a surrogate variable at a given split results in a similar node impurity measure (as would a competitor) but also mimics the chosen split itself in terms of which and how many observations go which way in the split.

Missing Values

As mentioned before, one of the advantages of classification and regression trees is their ability to accommodate missing values. If a response variable is missing, that observation can be excluded from the analysis, or, in the case of classification problem, treated as a new class (e.g., missing) to identify any potential patterns in the loss of information. If explanatory variables are missing, trees can use surrogate variables in their place to determine the split. Alternatively, an observation can be passed to the next node using a variable that is not missing for that observation.

Observation Weights

There are a number of instances where it might be desirable to give more weight to certain observations in the training set. Some examples include if the training sample has a disproportionate number of cases in certain classes or if the data were collected under a stratified design with one strata having greater or lesser sampling intensity. In these cases, observations can be weighted to reflect the importance each should bear.

Variable Importance

The importance of individual explanatory variables can be determined by measuring the proportion of variability accounted for by splits associated with each explanatory variable. Alternatively, one may address variable

importance by determining the effect of excluding variables in turn and by assessing the resulting predictive accuracy of the resulting models.

Model Assessment

While a single measure of error may be used to pick the optimum tree size, no single measure of error can capture the adequacy of the model for often diverse applications. Consequently, several measures of error may need to be reported on the final model. In classification problems, these may include the misclassification rate and kappa. Kappa measures the proportion of correctly classified units after accounting for the probability of chance agreement. In classification problems involving only a zero–one response, additional measures include sensitivity, specificity, receiver operating characteristic (ROC) curves with associated area under the curve (AUC). In regression problems measures of interest might include correlation coefficients, root mean squared error, average absolute error, bias, and the list continues. The literature on error assessment is vast. The point here is that an optimal tree size may be determined using one criterion, but often it is necessary to report several measures to assess the applicability of the model for different applications.

Enhancements through Ensemble Methods

While classification and regression trees are powerful methods in and of themselves, much work has been done in the data mining and machine learning fields to improve the predictive ability of these tools by combining separate tree models into what is often called a committee of experts, or ensemble. Following is a very brief description of some of these newer techniques using classification and regression trees as building blocks.

Bagging and Boosting

Two simple enhancements to tree-based methods are called bagging and boosting. These iterative schemes each produce a committee of expert tree models by resampling with replacement from the initial data set. Afterward, the expert tree models are averaged using a plurality voting scheme if the response is discrete, or simple averaging if the response is continuous. The difference between bagging and boosting is the way in which data are resampled. In the former, all observations have equal probability of entering the next bootstrap sample; in the latter, problematic observations (i.e., observations that have been frequently misclassified) have a higher probability of selection.

Random Forests

Another recent ensemble method is called 'random forests'. In this technique, a bootstrap sample of the training data is chosen. At the root node, a small random sample of explanatory variables is selected and the best split made using that limited set of variables. At each subsequent node, another small random sample of the explanatory variables is chosen, and the best split made. The tree continues to be grown in this fashion until it reaches the largest possible size, and is left unpruned. The whole process, starting with a new bootstrap sample, is repeated a large number of times. As in committee models, the final prediction is a (weighted) plurality vote or average from prediction of all the trees in the collection.

Stochastic Gradient Boosting

Yet another ensemble method is known as stochastic gradient boosting. In this technique, many small classification or regression trees are built sequentially from residual-like measures from the previous tree. At each iteration, a tree is built from a random subsample of the data set (selected without replacement), producing an incremental improvement in the model. Ultimately, all the small trees are stacked together as a weighted sum of terms. The overall model accuracy gets progressively better with each additional term.

Software

A wide variety of software packages are available for implementing classification and regression trees. The R part library and affiliated packages, part of the R public domain statistical software, is widely used. Popular commercial packages include Salford Systems CART, Rulequest's See5 and Cubist, tree-based models in S-Plus, to name a few.

See also: Data Mining.

Further Reading

Breiman L (2001) Random forests. *Machine Learning* 45: 5–32.
Breiman L, Friedman RA, Olshen RA, and Stone CG (1984) *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth.
Clark LA and Pregibon D (1992) Tree-based models. In: Chambers JM and Hastie TJ (eds.) *Statistical Models in S*, pp. 377–419. Pacific Grove, CA: Wadsworth and Brooks.
De'ath G and Fabricius KE (2000) Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.
Everitt BS and Hothorn T (2006) *A Handbook of Statistical Analyses using R*. Boca Raton, FL: Chapman and Hall/CRC.
Friedman JH (2002) Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38(4): 367–378.

Hastie T, Tibshirani R, and Friedman J (2001) *The Elements of Statistical Learning*. New York: Springer.
Murthy SK (1998) Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining and Knowledge Discovery* 2: 345–389.
Quinlan JR (1993) *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
Ripley BD (1996) *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.
Steinberg D and Colla P (1995) *CART: Tree-Structured Nonparametric Data Analysis*. San Diego, CA: Salford Systems.
Vayssieres MP, Plant RP, and Allen-Diaz BH (2000) Classification trees: An alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* 11: 679–694.
Venables WN and Ripley BD (1999) *Modern Applied Statistics with S-Plus*. New York: Springer.

Climate Change 1: Short-Term Dynamics

G Alexandrov, National Institute for Environmental Studies, Tsukuba, Japan

Introduction

Short-term dynamics of the global carbon cycle is closely related to the concept of climate system: the totality of the atmosphere, hydrosphere, biosphere, and their interactions. Human activities have been substantially increasing the concentrations of carbon dioxide and other greenhouse gases in the atmosphere and thus inducing potentially adverse changes in the climate system. This tendency has become of public concern that led to the United Nations Framework Convention on Climate Change (UNFCCC). This convention suggests protection of carbon pools, enhancement of carbon sinks, and reduction of emissions from carbon sources.

Carbon Pools

Carbon pool (or reservoir, or storage) is a system that has the capacity to accumulate or release carbon. The absolute quantity of carbon held within at a specified time is called carbon stock. Transfer of carbon from one carbon pool to another is called carbon flux. Transfer from the atmosphere to any other carbon pool is said to be carbon sequestration. The addition of carbon to a pool is referred to as uptake.

Carbon Sink

Carbon sink is a process or mechanism that removes carbon dioxide from the atmosphere. A given carbon pool can be a sink, during a given time interval, if carbon inflow exceeds carbon outflow.

Carbon Source

Carbon source is a process or mechanism that releases carbon dioxide to the atmosphere. A given carbon pool can be a source, during a given time interval, if carbon outflow exceeds carbon inflow.

Carbon Budget

The estimates of carbon stocks and carbon fluxes form the carbon budget, which is normally used as a kind of diagnostic tool in the studies of the short-term dynamics of the global carbon cycle.

Carbon Budget Components

The components of the global carbon budget may be subdivided into fossil and dynamic categories (**Figure 1**).

Fossil Components

The fossil components are naturally inert. The stock of fossil organic carbon and mineral carbonates (estimated at 65.5×10^6 PgC) is relatively constant and would not dramatically change within a century. The lithospheric part of the carbon cycle is very slow; all the fluxes are less than $1\,\mathrm{PgC\,yr^{-1}}$. For example, volcanic emissions are