Putting the CART Before the Horse
Using Classification Analysis to Determine Sample Design

Mary Sutter, Opinion Dynamics Corporation, Oakland, CA
Katherine Randazzo, KVD Research Consulting, San Diego, CA
Pamela Wellner, California Public Utilities Corporation, San Francisco, CA

ABSTRACT

A large-scale evaluation of California investor owned utility education and training programs contained a remarkably heterogeneous population from which we needed to cost-effectively sample to ultimately determine the effects of the entire program. We had none of the traditional parameters usually used to stratify populations undergoing energy efficiency evaluations (such as energy use, building type, etc.). Additionally, the population contained individuals who were included in multiple groups or treatments, thereby confounding the appropriate stratification of the available parameters. We did, however, have information from a first evaluation wave to inform the sample for the second wave. That data was quantitative in many formats (dichotomous, categorical, ordinal, and continuous).

We used a decision tree method called Classification and Regression Tree (CART) to determine the most efficient sampling strategy. CART is a non-parametric technique that can select from among a large set of categorical and continuous variables, those that individually, or in combination, best predict the outcome variable of interest by splitting the sample into progressively more parsimonious subgroups using multiple predictors.

The CART analysis indicated that a random sample was the most efficient design for this evaluation. The analysis also highlighted that the importance of participant characteristics supersedes class characteristics, though this is not to say that class content and method of delivery are unimportant. The general content of the class is important in predicting action, as is the level of perceived knowledge gained by the participant, particularly within specific type of participants and classes. For sampling, this method would be useful for identifying subgroups to investigate in more detail.

Introduction

A large-scale evaluation of California investor owned utility education and training programs contained a remarkably heterogeneous group from which we needed to cost-effectively sample to ultimately determine the energy savings effects of the entire program. Stratification can often increase the efficiency of sampling over the simple random sample. While this is typical of many assessments, we had none of the traditional parameters usually used to stratify populations undergoing energy efficiency evaluations (such as energy use, building type, etc.). Additionally, the population contained individuals who were included in multiple groups or treatments, thereby confounding the appropriate stratification of the available parameters. We did, however, have information from a first evaluation wave to inform the sample for the second wave. That data was quantitative in many formats (dichotomous, categorical, ordinal, and continuous), some of which lent themselves better than others to analysis that would help the sample design for the second wave of assessment.

To help with the sampling strategy, we chose to use a decision tree method called Classification and Regression Tree (CART). CART is a non-parametric technique that can select from among a large set of categorical and continuous variables, those that individually, or in combination, best predict the outcome variable of interest by splitting the sample into progressively more parsimonious subgroups using multiple predictors. In this paper, we provide an overview of the programs under evaluation, details about CART as well as the results of its application to the sample design. We also provide the results of using CART as a tool to help understand the importance of certain parameters in predicting desired outcomes.

Overview of Programs under Evaluation

The California Public Utilities Commission contracted with Opinion Dynamics and their team members to perform a multi-year evaluation of the indirect impacts[1] of four large education and training programs within California. The four Investor Owned Utilities (IOUs) in California implemented these programs. The focus of the evaluation was the Energy Centers within each of the IOU service territories (see Table 1.) Previous evaluations (Wirtshafter, 2005, Newcomb Anderson, 2003) looked at many of these Energy Centers, but were not targeted towards determining indirect impacts.

Table 1: Utility Programs and the Corresponding Energy Centers

Utility Program	Utility	Energy Center
Education and Training	Pacific Gas and Electric	Pacific Energy Center (PEC)
		Education and Training Center (ETC)
		Food Service Technology Center (FSTC)
Education, Training and Outreach	Southern California Edison	Agricultural Technology Application Center (AgTAC)
		Customer Technology Application Center (CTAC)
		Technology and Test Centers (TTC)
Education and Training	Southern California Gas	Energy Resource Center (SCG ERC)
CCSE Energy Resource Center Partnership	San Diego Gas and Electric	Energy Resource Center (SDERC)
		California Center for Sustainable Energy (CCSE)

While there were twelve research questions posed for the overall evaluation, seven are relevant to this sampling discussion as they show the different levels of information required for the evaluation:

1. What types of behavioral changes are occurring as a result of the activities of each Center?
2. How likely is the Center to induce behavioral change?
3. What is the incremental change in awareness of energy saving opportunities as a result of the Centers?
4. What percentages of participants were fed into resource programs, and which programs were promoted?
5. What behaviors did people who received education or training from the Centers undertake?
6. What are the net energy savings associated with behaviors undertaken by those who received education or training from the program?
7. What are the characteristics of the classes and activities that are most closely associated with net energy saving behavioral changes?

[1] An evaluation of indirect impacts is part of the California Evaluation Protocols. This type of evaluation seeks to determine the behavioral impacts of information, education, marketing or outreach programs for which energy savings are expected. The output of this evaluation is net changes in behaviors and energy savings.

As these questions show, there was a desire to understand and quantify the overall impact of the programs and of each Center. To understand the nature of the assessment challenges, one must know a bit more about each Center. Very briefly, the Centers covered a wide range of activities:

- Classes, Seminars and Workshops
- Customer-specific Training, Demonstrations and Consultations
- Lending Libraries
- Outreach Activities (facility tours, trade shows, industry events)
- Information Dissemination (Displays, Exhibits, Brochures, website)
- Energy Efficiency Technology Testing
- Unique characteristics such as:
  - Targeting agricultural customers
  - Emphasis on improving training effectiveness through Train the Trainer programs
  - Creation of food service test methods

There were over 1,000 classes taught and over 54,000 participants in an 18-month time period. (See Table 2)

Table 2: Energy Center Class and Participant Information (January 1, 2006 to June 30, 2007**)**

Energy Center	Total Number of Classes	Number of Unique Classes[2]	Number of Participants
ETC	417	91	7,951
PEC	216	94	8,811
CTAC	186	56	8,133
SCG ERC	182	72	17,980
FSTC	66	12	1,083
AgTAC	95	56	2,356
SDG&E ERC	61	39	5,960
TTC	28	7	675
CCSE	34	31	1,288
TOTAL	1,285	458	54,237

While the evaluation was appropriately funded, there was still a need to sample efficiently and to be prudent about the cost of data collection. Failing to sample efficiently could compromise our ability to estimate savings from the programs and Centers. Hence the need for a judicious and efficient sample design, such as a stratified sample, to support data collection.

Overall Evaluation Design

Because there was a multitude of needs for the evaluation and the best way to sample from the population to enable us to answer the research questions was uncertain, we chose a multi-stage evaluation design. This entailed data collection in two waves. The first wave of data collection covers the first 18 months of the program and the second wave will cover the second 18 months of the program. The results of the first wave of data collection from instructors and participants are the basis of the CART analysis to either identify homogeneous sampling groups or determine that a random sample

[2] Many Centers offered the same class multiple times during the evaluation period. This number reflects the number of unique classes offered by each Energy Center.

across all participants is the best approach. The analysis could enable us to efficiently sample from the second wave of data collection.

Wave 1 Outputs / Wave 2 Requirements

We originally planned to perform a linear regression using information from the first wave of evaluation, the output of which would show which parameters were of highest predictive value for the Wave 2 sample design. However, the difficulty of handling multiple individuals across several of the groups as well as our expectation that the data included in the regression would strongly deviate from a normal distribution (required for a robust use of the statistical model), as well as the potential for multi-collinearity, led us to consider other options. To enable us to meet the Wave 2 evaluation needs using the outputs of the Wave 1 analysis, we chose a non-parametric decision tree method called Classification and Regression Tree (CART).

Methodology - CART Description

CART (Breiman, et al., 1984) is a specific algorithm and software, belonging to a class of decision tree methodologies sometimes referred to as recursive partitioning methods. It is a non-parametric technique that can select from among a large set of categorical and continuous variables, regardless of their distributional characteristics, those that individually, or in combination, best predict the outcome variable of interest by splitting the sample into progressively more parsimonious subgroups using multiple predictors (or *splitters*, as they are called in CART). In this study we utilized this approach to identify the characteristics of the adult classes where participants were more likely to undertake action after the class. Five characteristics of the CART approach make it ideally suited to the type of data likely to be encountered in this type of study:

1. No variable used in a CART analysis, independent or dependent, is assumed to follow any specific type of statistical distribution, nor is independence of observations assumed.
2. CART has a built-in algorithm for dealing with missing values in a manner that doesn't eliminate observations (method is discussed below).
3. Outliers, multi-collinearity, heteroscedasticity or distributional error structures that so commonly plague parametric models do not affect CART.
4. CART detects and assesses interactions in the data.
5. CART effectively examines a large number of variables as potential predictors and can produce a parsimonious solution using only a few of these.

CART Analysis

CART analyses produce results that are typically displayed in an inverted tree shaped diagram referred to as a "classification tree." A unique language is used to describe classification trees. Figure 1 presents a hypothetical analysis and introduces this language.

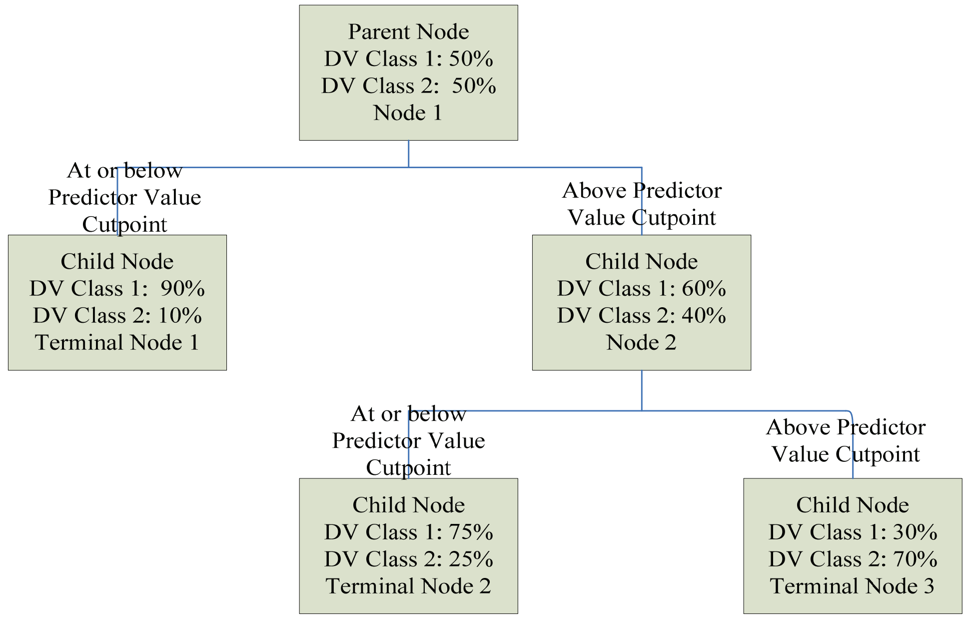

Figure 1. CART Output Example

The tree begins with a *parent node* (note that this term does not appear in trees based on actual data, but is understood to represent any node that results in further splits) that reports frequencies for each class of the dependent variable for the entire sample. In Figure 1, the dependent variable is represented as a binary outcome, with each class or category containing 50% of the observations. We utilize this example because it is relatively easy to follow. Note however that the dependent variable may represent a categorical (classification tree) or continuous (regression tree) outcome. The independent variables can be any combination of categorical or continuous variables; however, splitting will always result in two groups. In the case of categorical predictors, the categories will be divided into two groups, not necessarily consisting of contiguous categories. A continuous predictor, such as age, will be divided at the point along the continuum that best discriminates the two groups on the dependent variable using the specified splitting criteria. In Figure 1 the parent node is split into two *child nodes* (note that this phrase also doesn't appear in trees that show actual data, but it is understood to represent the two nodes that result from the splitting of a parent node). Each node produces a better classification outcome than is represented by its parent. In the case of the first child node in Figure 1, the 90% - 10% split cannot be improved and stopping criteria set within the software result in the termination of further splitting. In the case of the second child node, all variables are reassessed and a second split of the cases occurs. (Child) Node 2 becomes the parent node for the second split and the splitting criteria are again applied using all predictor variables. This means that the same predictor variable may enter the decision tree at more than one level, corresponding to its interaction with different variables. This represents a significant advantage of the CART approach and can illustrate the unique predictive power of the same variable at different levels of the tree. In this hypothetical example, the second split results in two terminal nodes and as all nodes are now terminal nodes the analysis is complete. Nodes are numbered systematically. Terminal nodes are numbered by one system and non-terminal nodes by another. Terminal nodes are numbered from 1 to the highest number starting at the left of the tree and moving counter clockwise. Non-terminal nodes are numbered starting at the top level of the tree moving down by level. Within a level, the numbering moves from left to right.

In addition to splitters, CART analyses identify two other types of potentially useful predictor variables known as *surrogates* and *competitors*. Surrogates are variables that act in a similar manner to the splitter (i.e., tending to place the same cases in the same nodes as the primary splitter), and are used to make the splitting decision in observations containing missing values for the splitter. Where there is no surrogate, all observations with a missing value on the splitter will be placed in the node with the majority of observations, and therefore available for the next split from that node. Competitors are variables that might purify nodes at almost the level of the selected primary splitters, but are not entered into the model because the selected variables were at least slightly stronger. They differ from surrogates in that they do not act similarly to the primary splitter. Thus a CART approach has the advantage of not only identifying the most efficient splitters, but also other potentially important predictors (i.e., surrogates and competitors).

In addition to producing classification and regression trees, CART analyses provide a list of predictor variables and their importance weights. The highest importance weight of 100 is assigned to the variable that, over the whole tree, best distinguishes groups and subgroups on the dependent variable. All other variables are assigned scores relative to 100 that represent their purification power relative to the most important variable. These weights can be described further in terms of whether the predictor variable was an efficient splitter, and can be reported considering or not considering surrogates.

The CART software allows several methods of cross-validation to provide a realistic estimate of the sample-specific effects in the tree. The method used in the present study is one that derives the main model from the whole sample (the *learning sample*), but generates predictive stability rates on test samples of successive withholdings of one-tenth segments of the sample. Ultimately 10 trees are generated on different sets of 9/10 of the sample, and are the basis of reported error rates for the test sample that can be compared to those from the learning sample.

For further details on CART and similar software, see Breiman et al. (1984); Steinberg and Colla (1995); Steinberg et al., (1998); Yohannes and Webb (1998); Zhang and Singer (1999); and Lemon et al., (2003).

The goal of the analysis was to create subgroups of participants that were homogeneous within each group and heterogeneous across groups so that the entire range could be sampled most efficiently. In this situation, obtaining "null" results, i.e., finding that CART could not identify any good splitters, would be an actionable outcome, since that would mean that a simple random sample would be the most efficient sampling design for wave two data collection.

Results - Application of CART to Wave 1 Outputs

The data output for Wave 1 consisted of information from two sources: a survey of all class instructors and the class participant surveys. Instructor surveys contained information regarding class structure such as length of class, percentage of time in each teaching format (i.e., lecture, hands-on exercises, etc.), and focus on energy efficiency (low, medium, high). Participant surveys contained demographic information and multiple questions regarding knowledge changes and actions taken. Six knowledge questions, each asking for response on a seven point scale, were averaged to create a knowledge index. Actions undertaken questions were summed (across multiple questions that were a Yes/No question, with Yes=1 and a No=0) and then converted to a percentage to create an action index for CART analysis. The number of questions regarding actions undertaken varied depending on whether the participant was an end user or market actor as the opportunities for actions to be taken vary. As such, the percentage was used to create a common metric across all participants. Additionally, the participant survey addressed the designated impact module by end use (i.e., lighting, HVAC, refrigeration, etc.). This variable served as a proxy for the focus of the class within the CART analysis.

Prior to analysis, the dependent variable or target variable of actions taken was collapsed into three levels: no actions, those who had action percentages from 1% to 74% of the total possible actions, and those who had 74% or higher.

Results of CART Analysis for Sampling

The second stage sample for this study is meant to serve several study goals, but the primary goal is to gather information from the sample participants in order to estimate the kWh savings achieved from the actions. While random sampling was still a consideration, CART was also considered as a means of identifying sample strata. The point of the CART analysis in this case was to identify variables that efficiently categorize participants into homogeneous groups on quantity of post-participation actions taken, using only variables that would be available before the next participant survey was fielded (i.e., instructor survey questions). Clearly, we could not sample based on variables we had not yet collected. Since the only variables that would be available at that time would be class descriptors, the initial trees used only those variables as potential splitters.

The resulting tree is shown in Figure 2. CART creates many trees but selects the one that is most efficient in node purity versus tree complexity, i.e., one with too many nodes. The most efficient tree by CART's standards had 6 terminal nodes and is presented in Figure 2.

The first level of analysis is to observe the rate of correct classification of cases. In this case, the no-action participants were the easiest to predict, and in the learning (full) sample, 48% were correctly classified by the six terminal nodes, and 42% of the time in the test sample[3]. The low-action participants (1-74%) were correctly classified 45% of the time in the learning sample, and 62% of the time in the test sample. Finally, those who took 75% or more actions were correctly classified 31% of the time in the learning sample and 24% in the test sample. The good news is that the results are relatively stable because the cross-validation procedure (test sample) resulted in very small changes in correct classification rates for the high and low categories. The bad news is that only two of the three groups were well predicted.

Another method of assessing the overall performance of the model is to evaluate the reduction in error from the initial node to the final (selected) tree. CART produces such an index which, in this case, was .886. This is the predictive accuracy compared to the initial node and is subtracted from 1 to yield the level of improvement due to the final tree splitters. Subtracting .886 from 1 produces .114 which translates to an 11.4% improvement. This is interpreted as an R^2 is interpreted, i.e., as explained variance. The formal effect size is the explained variance divided by the unexplained variance, and the result is called f^2. In those terms, this tree's (Figure 2) effect size is 0.13, which compares to the medium effect size, as defined by Cohen (1986), of 0.17. We can consider this tree useful in terms of content, but this is too inefficient to be used for sampling to estimate a population parameter.

The conclusion from this analysis was that simple random sampling would be the more efficient approach for the central research question of this study. However, the content of the trees are still of interest in thinking about what it is about classes and participants that impacts participant behaviors after taking the class. This will be addressed in the next section.

[3] Remember that the learning sample prediction is information from the entire sample while the test sample prediction is from multiple runs of 90 percent of the sample. When these values are similar, the predictions are considered to be stable.

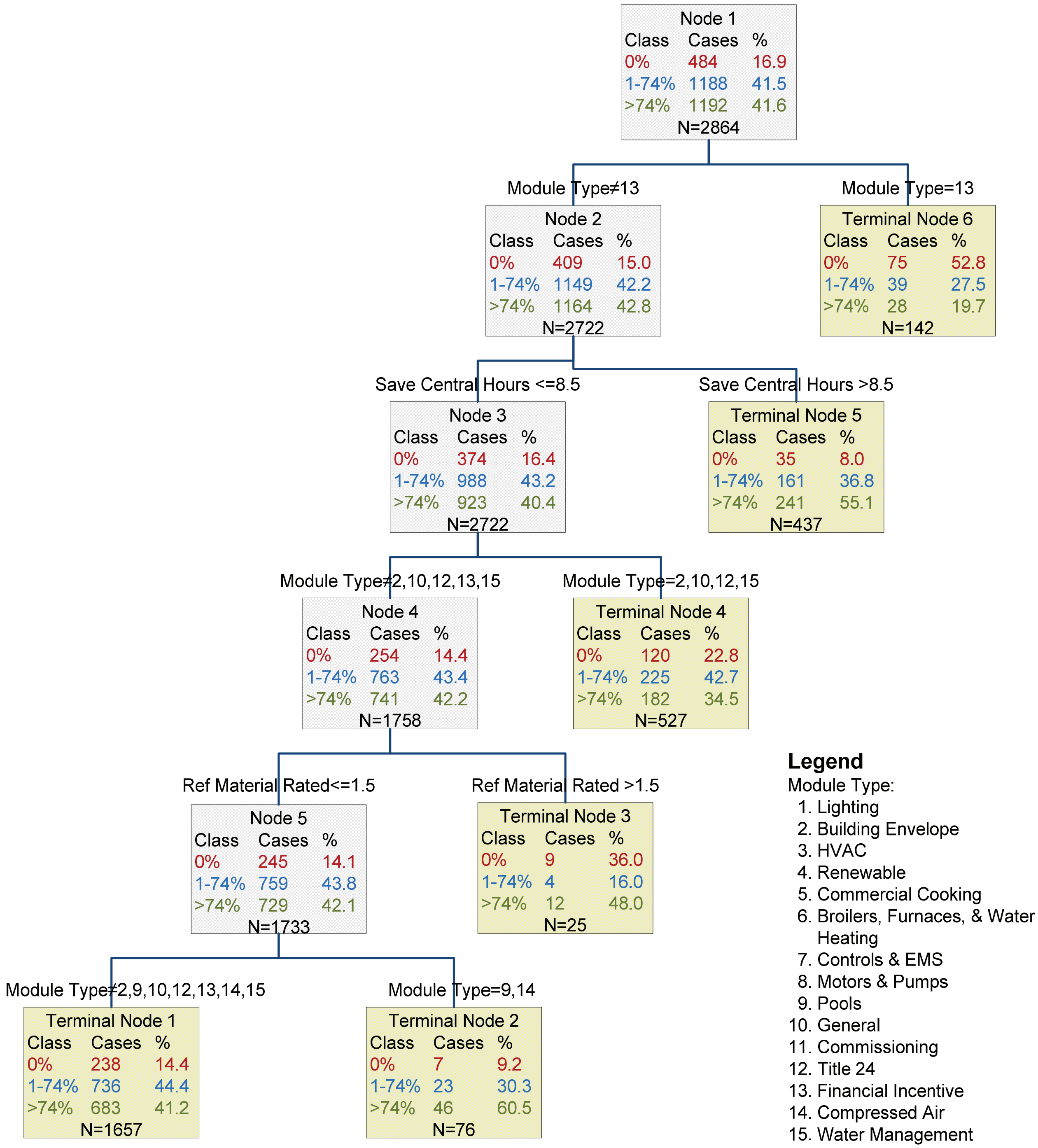

Figure 2. CART Results for Sampling – Class Descriptions Only

Results of CART Analysis for Predicting Outcomes

Class Variables. The CART analysis proved useful for identifying the most efficient sampling method, simple random sampling. Additionally, it helps analyze the drivers of participant actions considering only class-related variables. Analyzing Figure 2 in more detail shows a beginning or base node that contains all 2864 participants. Of those, 16.9% took no actions, 41.5% took some but less than three quarters of all possible actions noted in the survey, and 41.6% took three-quarters or more of the possible actions. Among these participants, those taking classes focusing on channeling participants into a resource acquisition program (based on the splitter Module Type called "financial incentives" Module Type=13 for the analysis) showed themselves less likely to undertake action than those of any other class type. As seen in Terminal Node 6 (TN6), the percentage of participants who took no action increased from 16.9% (in the parent node) to 52.8%. There are 142 participants in this node, and 75 of them took no action.

The larger group that took classes of any other type (the vast majority) looks very much like Node 1. They were, however, subject to a further split based on a splitter variable where energy savings is one of several themes addressed in the class and is a central component of the class, which further separated out the group (in TN5) who took more than 8.5 hours of classes. When people spent a large amount of time in classes where energy savings were a central component, they were more likely to undertake many actions (241 or 55.1%). For those who were in class less often, there was a small increase in those taking no action, or taking some actions, and a small decrease in those taking many actions. This group, however, was further split into two child nodes based on the type of class. Specifically, those who took classes that focused on the building envelope, Title 24, water management, or general topics (TN4) were slightly less likely to undertake 75% or more actions and more likely to undertake no actions.

The next split occurred based on a class designator. Ranging from 0 to 4, this variable stands for the number classes in which there were reference material handouts. Where reference materials were present 1.5 or more times, there was a high percent of actions taken (TN3). However, this split resulted in identifying only a small number of participants overall.

The final split was again based on the type of class (Module Type). Those taking Pool or Compressed Air classes (TN2) were substantially more likely to undertake 75% or more actions (60.5% did so). As with the previous split, though, this pulled out only a small number of participants (n=76). The last terminal split (TN1) is actually very similar to the original node (node 1) in terms of actions taken.

CART makes the splits noted above by selecting the variable that most efficiently splits the parent node into two child nodes. However, many times other variables would do almost as good a job as the one selected. Since they do not appear in the tree itself, they could be overlooked or assumed to be of no predictive value. Variables of this type are either surrogates or competitors. Surrogate variables act in a similar way to the primary splitter, i.e., it would assign the same cases to the same nodes, but less efficiently (which is why it is a surrogate and not a primary splitter). A competitor is a variable that may be almost as efficient as the primary splitter, but acts differently from the primary splitter (i.e., it will split the cases differently). Table 3 lists the tree splitters for each node, along with their surrogates and competitors. This information allows the reader (and analysts) to take note of other variables that also predict actions.

Table 3. Splitters and Their Surrogates & Competitors: Class Variables

Parent Node	Primary Splitter	Surrogates	Competitors
Node 1	Module type	Class provides examples	Energy savings is a central theme
		Energy savings is only theme	Uses hands-on exercises
Node 2	Energy savings is a central theme	Class aimed at residential sector	Module type
		Class delivery—lecture	Number of classes taken
		Class format—group class	
		Class aimed at trade professionals	
Node 3	Module type (end use)		Instructor provides demonstrations
			Rebates given brief discussion

Parent Node	Primary Splitter	Surrogates	Competitors
			Class delivery—Other
			Expect expertise of participant—basic
			Class provides examples
Node 4	Reference material provided		Module type
			Rebates given brief discussion
			Expect expertise of participant—basic
			Energy savings one of many themes
			Number of classes taken
Node 5	Module type		Rebates given brief discussion
			Instructor provides demonstrations
			Class delivery—video
			Class delivery—group discussion

In this tree, class characteristics that were also predictive, but operated only as surrogates (i.e., they were used where there were missing values on the primary splitter) included: class provides examples, energy savings is the only theme, class is aimed at residential sector, class delivery is lecture, class format is group class, and class aimed at trade professionals. There are more competitors than surrogates, and they are listed in Table 3. Clearly, the methods of delivery play a part as well as the general content and sector orientation of the classes.

All Variables. Since we have gone beyond using CART for identifying sample strata, we need not limit the analyses to class-level variables. We can take our understanding of what promotes post-class action to a deeper level by including information about the participants and what they experienced. Classification trees were developed using the action variable in its collapsed (trichotomy) form (Figure 3) as was done in the first tree.

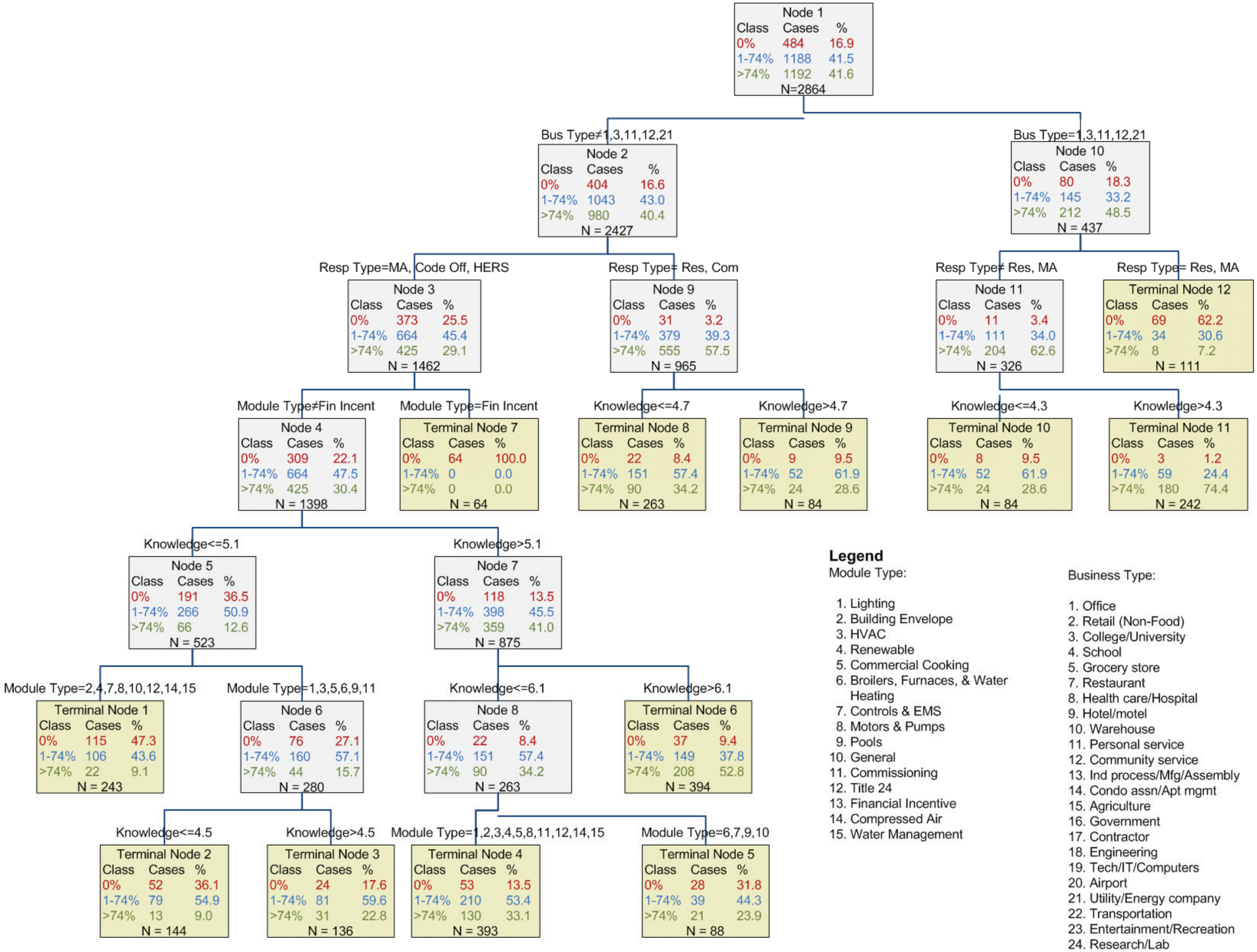

Figure 3. CART Results – All Variables

Figure 3 is the tree that CART chose as the optimal one, and it contains 12 terminal nodes. This tree shows about 35% improvement over the root node, which is a very large effect analogous to f^2 = 0.53, compared to Cohen's large f^2 effect size of 0.35. In terms of correct classification, those who took no action were correctly classified 68% of the time on the Learn sample, and 66% on the Cross-Validation (Test) sample; those who took 1% to 74% of the actions in the survey were correctly classified by this tree 42% of the time in the Learn sample, and 36% for the Test sample. Finally, those who took 75% or more of the actions were correctly classified 72% of the time and 67% for the Learn and Test samples, respectively. These figures indicate that: 1) high-action participants and no-action participants are best classified, and 2) the results are very stable since the Test and Learn sample results are very similar.

The starting distribution for the action levels are, of course, the same as in the Class variable tree. However, when there are new, individual participant-level variables available, the first splitter is different. The first splitter is the participant's business type and does not result in any terminal nodes, but does create two more parent notes of unequal size. Node 10 then creates a small terminal node (TN12) containing 111 participants that are residential end users or market actors working in one of the following categories: office, college/university, personal services, community services, or utilities, or energy companies. This group has a very high non-action rate. Of those 111 participants, 62.2% are in the lowest category. Those from all other types of participants (Node 11) have a very high level of action. However, they are further split by their self-reported change in knowledge into a very high action group (TN11) or a group with some action (TN10). Specifically, those who indicated a change in knowledge greater than 4.3 (on a 7 point scale) were split into a group of 242, of whom only a small percent failed to take action, while 74.4% took at least 75% of the possible actions open to them in the survey.

The respondent types on the other side of the original split (Node 2) had five more split levels before terminal nodes were found. These splits were also based on type of respondent, type of class, and knowledge gain. Probably the most interesting split was TN7, where all people ended up in the no-action level. These are the people who were in TN6 of the previous tree (Figure 2), but with some of them already removed from the analysis by this point.

The final splits originating from Node 2 are based on the knowledge variable for half and on the type of module for the other half. The knowledge level comes into play at different levels of knowledge gained for different parts of the tree. TN6 comes from a splitting value of 6.1, TN8 and TN9 result from a knowledge rating of 4.7, and TN2 and TN3 come from 4.5. This is an example of the strength of the CART analysis as it brings the knowledge variable into play at multiple times and in differing levels.

In addition to the information that has just been described that comes from the detail of the tree, two larger patterns are important. One is that participant characteristics predict actions far more than class characteristics. Only one class variable entered the tree as a primary splitter, and that was the class topic. The second pattern to note is more methodological, and highlights one advantage of CART over parametric methods: two variables entered the model two or more times, and different cut points or categories were optimal at different levels of the tree. These represent interaction effects that could be missed in a parametric model unless the analyst knew in advance what interactions to test for.

Conclusions

The CART method proved that a random sample would be more efficient than using stratum markers for the Wave 2 data collection and analysis. However, it would be very useful in identifying subgroups of participants to be studied in more depth to understand the reasons for undertaking action or non-action. One drawback of this method as a way to design samples is that the analysis requires a substantial amount of information as inputs into the model. Generally, this level of data is not available for helping to determine sample design prior to surveys. However, it was perfect for our use in this two-stage plan and we would recommend it when there is data to support its use.

The method proved useful in helping to understand more about interactions and characteristics than before. For instance, the importance of participant characteristics supersedes class characteristics, though this is not to say that class content and method of delivery are unimportant. Clearly, the general content of the class is important in predicting action, as is the level of perceived knowledge gained by the participant, particularly within specific types of participants and classes. Methods of class delivery entered the trees mainly as surrogates and competitors, indicating that these delivery methods are also important.

Differences among participant types could be fruitfully investigated toward understanding why some groups of people are highly productive following classes while others are not at all productive. Using CART to identify some of these subgroups could provide sampling efficiency in studying reasons for or barriers to action. Studies focusing on these groups could improve class content and delivery for them. The ability of CART to identify interactions can be very useful in these situations.

References

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

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. 2003. "Classification and regression tree analysis in public health: methodological review and comparison with logistic regression." *Annals of Behavioral Medicine*, 26, 172-181.

Newcomb Anderson Associates. 2003. *Measurement and Evaluation Study of PG&E's Local Crosscutting Pacific Energy Center Program.* December.

Steinberg, D., & Colla, P. 1995. *CART: Tree-structured non-parametric data analysis.* San Diego, CA: Salford Systems.

Steinberg, D., Colla, P., & Martin, K. 1998. *CART Classification and regression trees: Supplementary manual for Windows*. San Diego, CA: Salford Systems.

TecMarket Works Team. 2006. *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals.* April.

Wirtshafter Associates Inc, KEMA Consulting Inc, and Kreitler Research and Consulting. 2005. *Evaluation of the 2003 Statewide Education and Training Services Program.* June.

Yohannes, Y., & Webb, P. 1998. *Classification and regression trees: A user manual for identifying indicators of vulnerability to famine and chronic food insecurity.* International Food Policy Research Institute, Washington, D.C. Mimeo.

Zhang, H., & Singer, B. 1999. *Recursive Partitioning in the Health Sciences.* New York: Springer-Verlag.