Optimal Data Analysis
Vol. 5 (June 6, 2016), 58-61

2155-0182/10/$3.00

Pruning CTA Models to Maximize PAC

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

In CTA weighting by prior odds is used if a model is sought to maximize ESS, which is explicitly optimized by a pruning algorithm that deconstructs a fully-grown model into all nested sub-branches and then reassembles all possible combinations of sub-branches to identify the configuration with greatest ESS.[1] In contrast, unit-weighting is used if a CTA model is sought to maximize PAC, explicitly optimized using the pruning algorithm to reassembles all possible sub-branch combinations and identify the configuration with greatest PAC.

CTA-based models typically initiate with a root variable featuring two emanating branches that separate the model into left-hand and right-hand "sides" or "sections". In general, root variables with X (an integer greater than one) emanating branches separate a model into X sections: three branches create left-hand, middle and right-hand sections; four branches create outside and inside left- and right-hand sections; etcetera.

The algorithm that explicitly maximizes PAC of a CTA model in a given application is the same algorithm used to explicitly maximize ESS of a CTA model[1] except the objective function being maximized is PAC. Identifying the CTA model that explicitly maximizes overall PAC requires identifying every possible sub-branch for every branch emanating from the root variable.

For example, Figure 1 is the left-hand branch of a hypothetical CTA model with three nodes: A (root), B (middle node), and C (end node of the branch). Coded as L3 (L for left-hand, 3 for 3 nodes), this branch produces three model end-points (independent sample strata) with 10, 20, and 30 observations.

Figure 1: Three-Node Left-Hand Branch "L3" of Hypothetical CTA Model

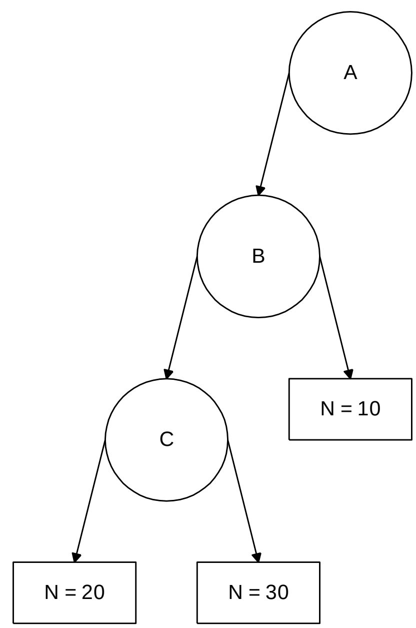
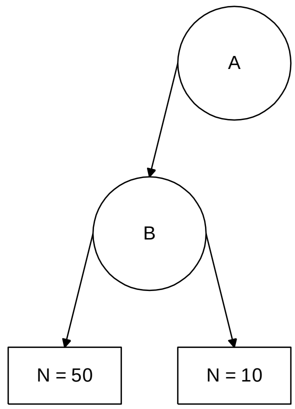


Figure 2 illustrates the left-hand branch with the end node C trimmed: observations in the two associated endpoints are aggregated into the left-hand endpoint of model L2: L for left-hand, 2 for 2 nodes.

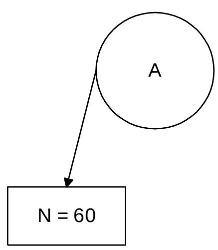
2155-0182/10/$3.00

Figure 2: Two-Node Left-Hand Sub-Branch "L2" of Hypothetical L3 Branch



Finally, Figure 3 is L2 with middle node B trimmed: observations in the two associated endpoints are aggregated into the left-hand endpoint of model L1: L for left-hand, 1 for 1 node.

Figure 3: One-Node Left-Hand Sub-Branch "L1" of Hypothetical L2 Branch



Imagine the right-hand branch of the hypothetical CTA model consisted of two nodes: A (the left- and right-hand sides share the root attribute) and D (at the end of the right-hand branch). The right branch involving two attributes (A, D) is coded R2 (R for right-hand, 2 for 2 nodes), and the trimmed right branch with one attribute (D collapsed into A), R1 (R for right-hand, 1 for 1 node).

Discussed in detail elsewhere[1], the first step of the pruning algorithm is obtaining a partial confusion table for all sub-branches of the model (L1, L2, L3, R1, and R2). The second step of the algorithm is obtaining a complete confusion table for every unique combination of left (L1-L3) and right (R1, R2) sub-branch. In this example there are six unique combinations: L1-R1; L2-R1; L3-R1; L1-R2; L2-R2; and L3-R2. The optimal CTA model for the application is selected as the combination of sub-branches with confusion table having maximum PAC.

Operationally, using the optimal pruning algorithm is simplified when every emanating branch except for one immediately terminates in a model endpoint. In the present example imagine the right-hand branch terminated in an endpoint, leaving only R1 on the right-hand side of the model. In this example there are thus three unique combinations (L1-R1; L2-R1; L3-R1) that must be compared. An example of such an application is presented below.

Predicting Maximum Patient Satisfaction

Data for this exposition were obtained from patients receiving care in the emergency department (ED) of a private Midwestern hospital who were mailed and returned a completed satisfaction survey.[2] Items were answered using a five-point Likert-type scale: 1=very poor, 2=poor, 3=fair, 4=good, 5=very good. The class variable was overall satisfaction with care received in the ED ("sat"): data of 671 patients rating overall satisfaction as good were compared with data of 1,045 patients rating overall satisfaction as very good. Potential attributes were patient ratings of six aspects of nurse care behavior (N1=courtesy; N2=took patient's problem seriously; N3= attention paid to patient; N4=concern to keep patient informed; N5=concern for patient privacy; N6= technical skill), and six aspects of physician care behavior (P1=waiting time in treatment area; P2=courtesy; P3= took the patient's problem seriously; P4=concern for patient comfort: P5=explanation of test/treatment; P6=explanation of illness/injury).

Data were analyzed via hierarchically optimal CTA conducted using CTA software.[1] Statistical power analysis indicated a minimum endpoint of N=100 observations yields 90.4% power to test the *a priori* hypothesis that higher

2155-0182/10/$3.00

performance ratings predict higher satisfaction ratings with one-tailed $p<0.01$ for a moderate effect (Cohen's $d=0.5$).[1] To explicitly maximize PAC, weighting by prior odds is turned off[1]:

```
OUTPUT satis.out;
OPEN satis.dat;
VARS sat n1 TO n6 p1 TO p6;
CLASS sat;
ATTR n1 TO n6 p1 TO p6;
MISSING all (-9);
PRIORS OFF;
MC ITER 10000 CUTOFF .05 STOP 99.9;
MINDENOM 100;
PRUNE .05;
GO;
```

The resulting HO-CTA model is given in Figure 4 (PAC was stable in LOO analysis[1]).

Figure 4: Five-Strata L4-R1 CTA Model

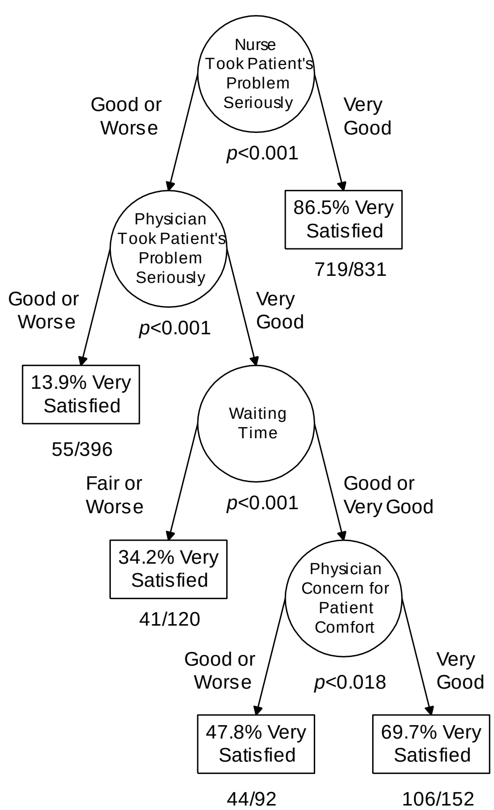


Table 1 presents the confusion table[1] for the L4-R1 CTA model illustrated in Figure 4.

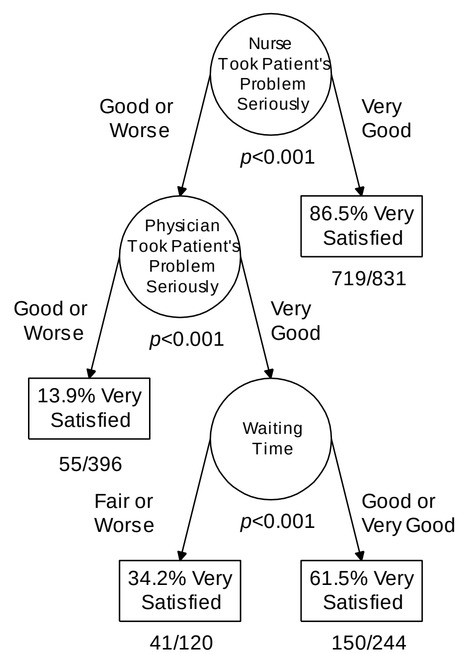
Table 1: Confusion Table for L4-R1 CTA Model

	Predicted Satisfaction	
Actual	Good	Very Good
Good	468	158
Very Good	140	825

For the L4-R1 model, overall PAC = (468+825) / (468+158+140+825), or 81.3%.[1]

The L4-R1 model must now be pruned to ensure explicit maximization of PAC (current CTA software prunes CTA models to explicitly maximize ESS, but not PAC). Retracting the last node in the L4-R1 model yields the simpler L3-R1 model illustrated in Figure 5.

Figure 4: Four-Strata L3-R1 CTA Model



This model yielded the classification performance summarized in Table 2.

2155-0182/10/$3.00

Table 2: Confusion Table for
L3-R1 CTA Model

	Predicted Satisfaction	
Actual	Good	Very Good
Good	420	206
Very Good	96	869

For the L3-R1 model, overall PAC = (420+869) / (420+206+96+869), or 81.0%. For L2-R1 PAC = 78.5%, and for L1-R1 PAC = 77.3%. It has thus been determined that the L4-R1 CTA model yields the greatest PAC in this application, in both training and generalizability (LOO) analysis.

Systematic research on models that explicitly maximize PAC is just beginning, but early findings suggest maximum PAC models are often relatively stable in analyses assessing potential cross-generalizability. Furthermore, while pruning to maximize ESS has been found to sometimes greatly affect the nature of the maximum PAC model that is identified, this procedure has also produced anomalous affects such as the pruned model having lower ESS than the non-pruned model! This is because maximization of overall PAC (that isn't normed against chance) and maximization of ESS (that is normed against chance) are different and independent enterprises: tools that are used to maximize ESS shouldn't be used to alter models that by design maximize PAC, and the opposite is also true. Stated another way, weighting by prior odds is disengaged to identify a maximum PAC model, so what is the statistical motivation to use a pruning procedure that reengages the use of weighting by priors to alter the model so as to maximize ESS? This disconnect between PAC and ESS methods is particularly quirky in applications for which an attribute has $p < 0.05$ when computed for PAC, but $p > 0.05$ when computed for ESS, and vice versa.

Considered from an "action" (logistic or operational) perspective the model is complex, and examination of the five-strata model fails to suggest any obvious avenues for meaningful future gains (i.e., qualitatively assessed as reflecting strong improvement) in PAC. Models that maximize ESS for the identical data have been much more parsimonious, and produced stronger effects when gauged against chance.[1]

References

[1]Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago: ODA Books.

[2]Yarnold PR (2014). What most satisfies Emergency Department patients? *Optimal Data Analysis*, *3*, 98-101.

Author Notes

The study analyzed de-individuated data and was exempt from Institutional Review Board review. No conflict of interest was reported.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., #163
Chicago, IL 60646
USA