Optimal Data Analysis
Vol. 8 (January 21, 2019), 26-29

2155-0182/10/$3.00

Maximizing Classification Accuracy of CART® Recursive Partitioning Tree Models Using Optimal Pruning

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

A classification tree pruning methodology that maximizes effect strength for sensitivity is demonstrated for a model which was developed using CART® software to predict influenza among primary care patients.

Most published empirical classification tree models initiate with a root variable having two emanating branches which separate the sample into left- and right-hand partitions.[1] Extension to applications having more initiating branches is straightforward: for example, three branches create left-hand, middle, and right-hand partitions. For clarity of exposition this paper considers models having two emanating branches.

Identifying the tree model explicitly maximizing mean sensitivity—and thus effect strength for sensitivity[2] (ESS)—necessitates identifying every sub-branch of each emanating branch. For example, imagine a left-hand branch consisting of three nodes: A (root), B (middle attribute), and C (end of branch). This branch has two nested sub-branches: one involves only nodes A and B (C collapsed into B), and one involves only node A (C and B collapsed into A). This paper refers to the *left* branch having *three* nodes (A, B, C) as "L3"; to the trimmed *left* branch with *two* nodes (A, C collapsed into B) as "L2"; and to the trimmed *left* branch with *one* node (C and B collapsed into A) as "L1".

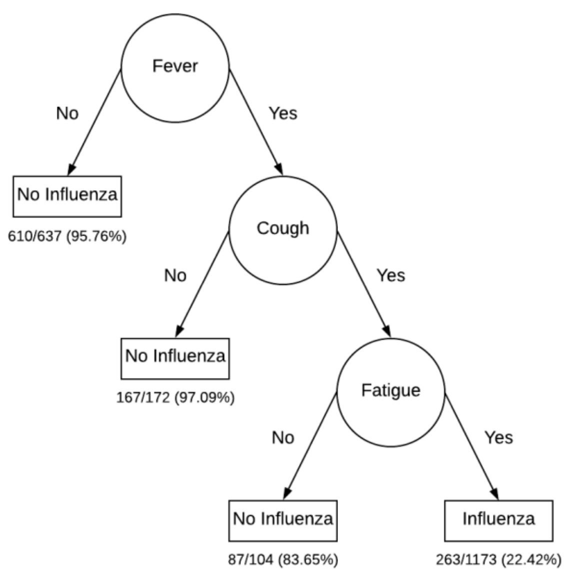
Further imagine this hypothetical tree model has a right-hand branch consisting of two nodes: A (both sides have the same root attribute) and D (end of the branch). The *right* branch having *two* attributes (A, D) is called "R2", and the trimmed *right* branch having *one* attribute (D collapsed into A) is called "R1".

Optimal pruning requires obtaining a confusion table—in which rows are actual class category and columns are predicted class category—for every unique combination of left and right (sub)branch. In the example the six unique combinations are L1-R1, L2-R1, L3-R1, L1-R2, L2-R2 and L3-R2. The optimized model is the combination of (sub)branches with associated confusion table yielding maximum ESS.[3]

Maximizing ESS of a Classification Tree Model Predicting Presence *vs*. Absence of Influenza, Created using CART® Software

A recent study[4] noted: "Rapid, cost-effective methods for diagnosing influenza are needed to enable appropriate prescribing. Multi-viral respiratory panels using reverse transcription polymerase chain reaction (PCR) assays to diagnose influenza are accurate but expensive and more time-consuming than low sensitivity rapid influenza tests. Influenza clinical decision algorithms are both rapid and inexpensive, but most are based on regression analyses that do not account for higher order interactions." Accordingly, CART (classification and regression trees) was used to estimate probability of influenza for a training sample of 2,087 enrollees presenting at ambulatory care centers for treatment of acute respiratory illness, and the resulting model was evaluated using a hold-out validity sample of 2,086 enrollees (Figure 1).

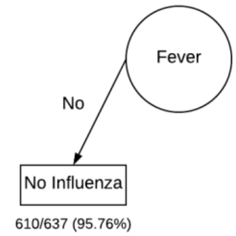
Figure 1: CART Model of Presence *vs*. Absence of Influenza: Hold-Out Validity Sample[4]



As seen, the root node of the non-pruned CART model has two emanating branches: the right-hand side has two additional nodes, and the left-hand side has no additional nodes.

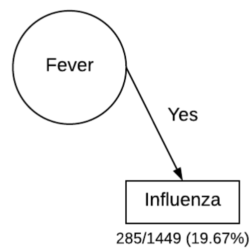
Figure 2 presents a schematic illustration of L1 and its associated confusion table, and Figures 3A-3C present schematic illustrations of R1-R3 and their associated confusion tables.

Figure 2: L1 Sub-Branch and Confusion Table



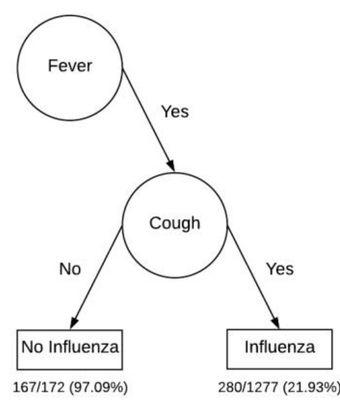
	Predicted	
Actual	No Influenza	Influenza
No Influenza	610	0
Influenza	27	0

Figure 3A: R1 Sub-Branch and Confusion Table



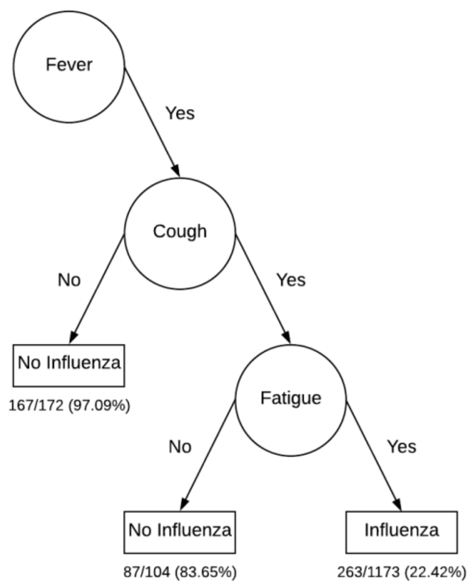
	Predicted	
Actual	No Influenza	Influenza
No Influenza	0	1164
Influenza	0	285

Figure 3B: R2 Sub-Branch and Confusion Table



	Predicted	
Actual	No Influenza	Influenza
No Influenza	167	997
Influenza	5	280

Figure 3C: R3 Sub-Branch and Confusion Table



	Predicted	
Actual	No Influenza	Influenza
No Influenza	254	910
Influenza	22	263

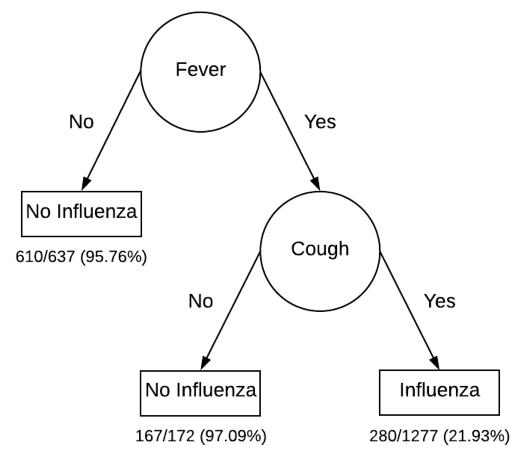
Table 1 gives integrated confusion tables for all four possible combinations of left (L1) and right (R1-R3) sub-branches, and their ESS.

Table 1: Classification Results for Every Combination of Left (L1) and Right (R1-R3) Sub-Branch

Model	Confusion Table	
L1-R1	Predicted	
Actual	No Influenza	Influenza
No Influenza	610	1164
Influenza	27	285
		ESS=25.74
L1-R2	Predicted	
Actual	No Influenza	Influenza
No Influenza	777	997
Influenza	32	280
		ESS=33.54
L1-R3	Predicted	
Actual	No Influenza	Influenza
No Influenza	864	910
Influenza	49	263
		ESS=32.99

The combination L1-R2 (Figure 4) has the greatest mean sensitivity (66.77%), yielding optimized ESS=33.54—corresponding to an effect of moderate strength.[2]

Figure 4: CART Model Having Maximum ESS



By definition the model which achieves maximum ESS performs best when compared against chance—and thus is the preferred model for *translational* application.[2] However, highest ESS does not imply that the model is best when considered from a *theoretical* perspective which includes the criterion of parsimony—this is the purpose of the D (distance) statistic.[5,6]

Minimizing D of a Classification Tree Model

Discovered after optimal pruning methodology, D norms ESS for the number of attributes in the

2155-0182/10/$3.00

model, indicating the number of additional effects having equivalent mean ESS which are needed to obtain a perfectly accurate model.[1,6] Presently, for L1-R1, D=5.77; for L1-R2, D= 5.94; and for L1-R3, D=8.12. Because the L1-R1 model has the *lowest* D statistic of all three CART models, L1-R1 is closest to representing a theoretically optimal solution. Although using "cough" and "fatigue" as attributes in the model improves empirical performance, these variables nevertheless degrade theoretical representation of influenza infection: attributes not presently assessed are thus needed to advance theoretical understanding in this realm.

Comment

It isn't surprising to find, yet again, that ODA methodology improves the classification and discrimination performance of models created using suboptimal legacy statistical methods—this has been demonstrated for an enormous variety of such (non)parametric methods.[5,7] It is axiomatic that multiattribute procedures which incorporate suboptimal methodologies in their formulation must also therefore be suboptimal. Compared *vs*. legacy approaches, classification tree methods developed in the ODA paradigm *cannot* be surpassed with respect to ESS (accuracy normed against chance), D (ESS normed for parsimony), control of Type I error within multiattribute models, variety of cross-validation methods which may be used, and demonstrated replication across samples and time.[5]

References

[1]Yarnold PR (2019). The structure of *perfect* optimal models with a two-category class variable and four or fewer endpoints. *Optimal Data Analysis*, *8*, 21-25.

[2]Yarnold PR (2017). What is optimal data analysis? *Optimal Data Analysis*, *6*, 26-42.

[3]Yarnold PR, Soltysik RC (2010). Maximizing the accuracy of classification trees by optimal pruning. *Optimal Data Analysis*, *1*, 23-29.

[4]Zimmerman RK, Balasubramani GK, Nowalk MP, Eng H, Urbanski L, Jackson ML, Jackson LA, McLean HQ, Belongia EA, Monto AS, Malosh RE, Gaglani M, Clipper L, Flannery B, Wisniewski SR (2016). Classification and regression tree (CART) analysis to predict influenza in primary care patients. *BMC Infectious Diseases*, *16*: 503. https://doi.org/10.1186/s12879-016-1839-x

[5]Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

[6]Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, *5*, 171-174.

[7]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC, APA Books.

Author Notes

No conflict of interest was reported.