Optimal Data Analysis
Vol. 7 (September 2, 2018), 59-65

2155-0182/10/$3.00

Identifying Maximum-Accuracy Cut-Points for Diagnostic Indexes via ODA

Ariel Linden, Dr.P.H. and Paul R. Yarnold, Ph.D.
Linden Consulting Group, LLC Optimal Data Analysis LLC

Maximizing the discriminatory accuracy of a diagnostic or screening test is paramount to correctly identifying individuals with *vs*. without the disease or disease marker. In this paper we demonstrate the use of ODA to identify the optimal cut-point which best discriminates between those with *vs*. without the disease (or marker) under study, for any diagnostic test. We illustrate this methodology using a dataset composed of a range of repeated biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). A logistic regression model was used to estimate the probability that each observation was from a person with *vs*. without PD as a function of 22 voice measurement variables, entered in the model as main effects only. Five different methods for computing a diagnostic cut-point on estimated probability are compared.

Among a smorgasbord of statistical procedures used in an effort to construct "accurate" tests, only "optimal" maximum-accuracy statistical methods—ODA[1-4], MultiODA[4-7], CTA[8-10] and novometrics[10-12]—are specifically formulated to explicitly identify models which maximize accurate classification. Model accuracy in the optimal paradigm may be operationalized in terms of percent (of total N) accurate classification (PAC; 0=no observations correctly classified, 50=N/2 observations correctly classified, 100= all observations correctly classified), or in terms of effect strength for sensitivity (ESS; 0=mean sensitivity expected by chance; ≤25= relatively weak effect; ≤50=moderate effect; ≤75=relatively strong effect; ≤90= strong effect; >90= very strong effect).[4,12]

In a comprehensive series of papers, the ODA and CTA frameworks were applied to observational data, and to data from randomized controlled trials, to draw causal inferences about treatment effects.[13-31] Herein we demonstrate how ODA is used to determine the optimal cut-point to maximally discriminate between those with *vs*. without disease, for any continuum of test values. Individuals are classified as disease positive if their test value surpasses the criterion cut-point, and disease negative if their test value falls below the cut-point. Model sensitivity is the proportion of true disease positives (individuals correctly predicted by the test as having the disease), and model specificity is the proportion of true disease negatives (individuals correctly predicted as not having the disease).[32]

2155-0182/10/$3.00

Receiver Operating Characteristic or ROC analysis is a popular approach for assessing and displaying the overall discriminatory accuracy of diagnostic tests, which involves plotting sensitivity vs. 1-specificity across the full range of values for the sample. Area under the curve (AUC) is then computed to assess the test's overall discriminatory ability. A test with perfect discriminatory ability has AUC=1.0, and a test unable to distinguish between individuals with *vs*. without disease has AUC=0.50.[32]

Using values computed in ROC analysis, several approaches have been developed to find the "optimal" cut-point on a diagnostic test at which individuals are "best" classified. The "nearest to (0,1) method" finds the cut-point on the ROC curve closest to the upper left-hand corner (i.e. the point with perfect sensitivity and specificity).[33] The "concordance probability method" defines the optimal cut-point as the point maximizing the product of sensitivity and specificity.[34] The "Youden index (*J*) method" defines the optimal cut-point as the point maximizing the difference between the true positive rate and false positive rate over all possible cut-point values.[35] The "Index of union method" defines the optimal cut-point as the point which minimizes the sum of absolute values of the differences between AUC and sensitivity and specificity, provided that the difference between sensitivity and specificity is minimum.[36]

We compare these methods to the "ODA method"[37-39] which defines the optimal cut-point as the point maximizing the average sensitivity and specificity over all possible cut-points, and yields the maximum possible effect strength for sensitivity (ESS) for the sample: ESS=0 indicates the accuracy expected by chance, and ESS=100 indicates perfect discrimination.[4,12]

Methods

Data

The dataset used presently is composed of 22 repeated biomedical voice measurement variables taken from 31 people, 23 with Parkinson's disease (PD), for a total of 195 observations with no missing data. The dataset is available at: https://archive.ics.uci.edu/ml/datasets/Parkinsons (see Little et al.[40] for details concerning the voice measurement variables).

Analyses

A logistic regression model estimated the probability that each observation was a person with *vs*. without PD as a function of the 22 voice measurement variables, entered in the model as main effects only.

An ROC analysis was then conducted in which true disease status (positive or negative for PD) was set as the reference variable and the predicted probabilities from the logistic regression were set as the classification variable. Optimal cut-points were computed using methods previously described.[33-36]

An ODA model[4,12] was generated in which the binary indicator for disease status was specified as the class variable, and the predicted probabilities variable from the logistic regression was specified as the attribute. A total of 25,000 Monte Carlo simulations were used to compute *P* values, and leave-one-out (LOO) analysis was conducted to assess cross-generalizability. LOO analysis is inherently one-tailed (directional), and is only conducted for models having $P \leq 0.05$ in training (total sample) analysis.[4,12]

Logistic regression and ROC analyses were conducted using Stata statistical software version 15.1 (StataCorp, College Station, Texas) and ODA models were generated using the ODA software package.[4]

Results

Figure 1 presents the ROC curve with markers indicating the cut-points developed using the different methods. The overall AUC statistic was 0.9520 which indicates that the model has overall high discriminatory ability on the full

2155-0182/10/$3.00

(training) sample. ODA and the Youden method produced identical cut-points at 0.90345, the nearest (0,1) method produced a cut-point at 0.74775, and the Liu and Unal methods produced identical cut-points at 0.88355.

The ODA and Youden methods both identified the optimal cut-point at which the sensitivity was 0.7415 and specificity was 1.0, yielding AUC at the cut-point of 87.07. This level of classification accuracy corresponds to ESS=74.15, indicating a relatively strong effect (defined as $50 \leq ESS < 75$).[4,12]

The nearest (0,1) approach identified the optimal cut-point with sensitivity=0.8776 and specificity=0.8333, yielding AUC at the cut-point of 0.8554, corresponding to a relatively strong ESS=71.09.

Both the Liu and Unal methods identified the optimal point with sensitivity=0.7823 and specificity=0.8333, yielding AUC at the cut-point of 87.03, and relatively strong ESS=61.56.

In contrast to other methods, ODA also computes an exact (permutation) *P*-value to assess statistical significance of the classification performance obtained using the cut-point in the full sample and in one-sample jackknife ("leave-one-out" or LOO) analysis. The *P*-value<0.0001 for both full sample and LOO analysis indicates that the ODA optimal cut-point was a statistically significant, generalizable discriminator of individuals with vs. without PD.

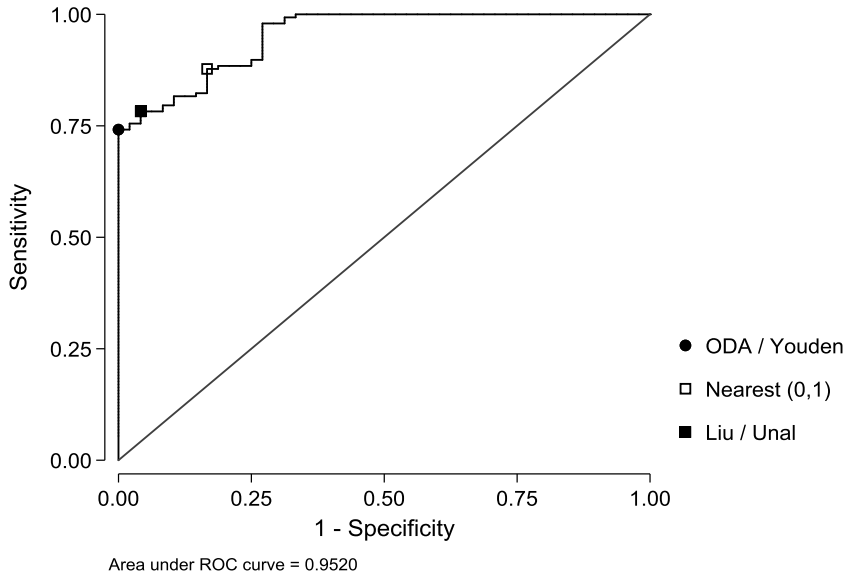

Figure 1: ROC curve highlighting the optimal cut-points computed using various methods

Discussion

In this paper we have demonstrated how ODA can be used to identify the optimal cut-point on a diagnostic test to maximally discriminate between those with and without the disease. While the various methods implemented here identify optimal cut-points using different criteria, the AUC at the cut-point can serve as an arbiter of discriminatory ability.

In this example, ODA and Youden's method identified cut-points which elicited

2155-0182/10/$3.00

identical AUC values which were higher than those of the other methods. The ODA and Youden methods will always concur *for applications involving two class categories*. The optimal ODA cut-point for such designs maximizes ESS which is computed as {[(sensitivity + specificity)/2] / .5}, and the optimal Youden cut-point for such designs maximizes *J*, computed as [(sensitivity + specificity) – 1]: a bit of algebra shows these definitions are isomorphic.

Another important issue is that in the present paper we purposely used logistic regression for generating the probabilities of having the disease, conditional on the 22 covariates to illustrate how ODA can be used in conjunction with conventional parametric models. However, we strongly advocate using CTA for identifying the optimal multivariable solution for classification problems (Linden & Yarnold[13,28]). ODA (and by extension CTA) should be considered the preferred approach over commonly-used parametric models because ODA avoids the assumptions required of parametric models (e.g. linearity, sufficient sample size, independence, etc.), provides exact (permutation) P-values, is insensitive to skewed data or outliers, and can handle any variable metric including categorical, Likert-type integer, and real number measurement scales.[4,12] And, as demonstrated here, unlike regression models ODA identifies optimal (maximum-accuracy) cut-points for variables of interest, facilitating the use of measures of predictive accuracy.

Furthermore, ODA has the capability to use cross-validation methods such as LOO which was used presently, in addition to hold-out, multiple-sample, test-retest, bootstrap[41] and other cross-validation methods to assess the generalizability of the model[4] to other individuals (outside of the current study sample) with similar characteristics.[42] Again, there is no equivalent in the parametric model-based framework, failing to provide insight into the likelihood that the findings would generalize.

References

1. Yarnold PR, Soltysik RC. Theoretical distributions of optima for univariate discrimination of random data. *Decision Sciences* 1991;22:739-752.

2. Soltysik RC, Yarnold PR. Univariable optimal discriminant analysis: One-tailed hypotheses. *Educational and Psychological Measurement* 1994;54:646-653.

3. Carmony L, Yarnold PR, Naeymi-Rad F. One-tailed Type I error rates for balanced two-category UniODA with a random ordered attribute. *Annals of Operations Research* 1998;74:223-238.

4. Yarnold PR, Soltysik RC. *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books, 2005.

5. Soltysik RC, Yarnold PR. The Warmack-Gonzalez algorithm for linear two-category multivariable optimal discriminant analysis. *Computers and Operations Research* 1994;21:735-745.

6. Yarnold PR, Soltysik RC, McCormick WC, Burns R, Lin EHB, Bush T, Martin GJ. Application of multivariable optimal discriminant analysis in general internal medicine. *Journal of General Internal Medicine* 1995;10:601-606.

7. Yarnold PR, Soltysik RC, Lefevre F, Martin GJ. Predicting in-hospital mortality of patients receiving cardiopulmonary resuscitation: Unit-weighted MultiODA for binary data. *Statistics in Medicine* 1998;17:2405-2414.

2155-0182/10/$3.00

8. Yarnold PR. Discriminating geriatric and non-geriatric patients using functional status information: An example of classification tree analysis via UniODA. *Educational and Psychological Measurement* 1996;56:656-667.

9. Yarnold PR, Soltysik RC, Bennett CL. Predicting in-hospital mortality of patients with AIDS-related *Pneumocystis carinii* pneumonia: An example of hierarchically optimal classification tree analysis. *Statistics in Medicine* 1997;16:1451-1463.

10. Yarnold PR, Soltysik RC. Globally optimal statistical classification models, I: Binary class variable, one ordered attribute. *Optimal Data Analysis* 2014;3:55-77.

11. Yarnold PR, Soltysik RC. Globally optimal statistical classification models, II: Unrestricted class variable, two or more attributes. *Optimal Data Analysis* 2014;3:78-84.

12. Yarnold PR, Soltysik RC. *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books, 2016. DOI: 10.13140/RG.2.1.1368.3286

13. Linden A, Yarnold PR. Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice* 2016;22:839-847.

14. Linden A, Yarnold PR. Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice* 2016;22:848-854.

15. Linden A, Yarnold PR. Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice* 2016;22:855-859.

16. Linden A, Yarnold PR, Nallomothu BK. Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice* 2016;22:860-867.

17. Yarnold PR, Linden A. Using machine learning to model dose-response relationships via ODA: eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis* 2016(5):41-52.

18. Linden A, Yarnold PR. Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice* 2016;22:868-874.

19. Linden A, Yarnold PR. Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice* 2016;22:875-885.

20. Yarnold PR, Linden A. Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis. *Optimal Data Analysis* 2016;22:65-73.

21. Yarnold PR, Linden A. Theoretical aspects of the D statistic. *Optimal Data Analysis* 2016;22:171-174.

22. Linden A, Yarnold PR. Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice* 2017;23:703-712.

23. Yarnold PR, Linden A. Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis* 2017;6:43-46.

24. Linden A, Yarnold PR. Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice* 2017;23:1299-1308.

2155-0182/10/$3.00

25. Linden A, Yarnold PR. Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice* 2017;23:1309-1315.

26. Linden A, Yarnold PR. The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis* 2018;7:28-35.

27. Linden A, Yarnold PR. Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice* 2018;24:353-361.

28. Linden A, Yarnold PR. Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice* 2018;24:380-387.

29. Linden A, Yarnold PR. Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice* 2018;24;740-744.

30. Linden A, Yarnold PR. Using ODA in the evaluation of randomized controlled trials. *Optimal Data Analysis* 2018;7:46-49.

31. Linden A, Yarnold PR. Using ODA in the evaluation of randomized controlled trials: application to survival outcomes. *Optimal Data Analysis* 2018;7:50-53.

32. Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. *Journal of Evaluation in Clinical Practice* 2006;12(2):132-139.

33. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cut-points using two ROC based criteria. *American Journal of Epidemiology* 2006;163:670–675.

34. Liu X. Classification accuracy and cut point selection. *Statistics in Medicine* 2012;31:2676–2686.

35. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–35.

36. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Computational and Mathematical Methods in Medicine*. 2017.

37. Yarnold PR, Soltysik RC (1991). Refining two-group multivariable classification models using univariate optimal discriminant analysis. *Decision Sciences, 22*, 1158-1164.

38. Yarnold PR, Hart LA, Soltysik RC (1994). Optimizing the classification performance of logistic regression and Fisher's discriminant analyses. *Educational and Psychological Measurement, 54*, 73-85.

39. Yarnold PR. UniODA vs. ROC Analysis: Computing the "Optimal" Cut-Point. *Optimal Data Analysis* 2014;3:117-120.

40. Little MA, McSharry PE, Roberts SJ, Costello DA, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. Biomedical engineering online. 2007 Dec;6(1):23.

41. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: An introduction to the bootstrap technique. *Disease Management and Health Outcomes* 2005;13:159-167.

2155-0182/10/$3.00

42. Linden A, Adams J, Roberts N. The generalizability of disease management program results: getting from here to there. *Managed Care Interface* 2004;17:38-45.

Author Notes

No conflict of interest was reported by either author.