Optimal Data Analysis
Vol. 9 (July 3, 2020), 207-208

2155-0182/10/$3.00

Status of Current MultiODA Research in the Optimal Data Analysis Laboratory

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

This note updates current MultiODA research activity in the ODA lab.

ODA software systems designed for 386-class PCs and 1990's main/supercomputers were used to study alternative linear multivariable optimal discriminant analysis or *MultiODA* formulations.[1-5] MultiODA identifies the linear model which explicitly maximizes overall *percent of accurately classified* (PAC) observations for a particular sample. For some applications it is logical to maximize overall PAC: for example, modeling used to encrypt an electronic message for efficient deep-space transmission, thus conserving satellite electrical power. And, MIP45 may also be instrumental in developing calculus to model finite stochastic processes.[6] It therefore seems natural to wonder why classification tree analysis (CTA) is the multiattribute statistical algorithm which is currently the most studied in the ODA laboratory.[7-11]

The primary reason is that the only ODA objective function adjusted to remove the effect of chance is the effect strength for sensitivity or ESS measure of classification accuracy—which CTA best maximizes.[12,13] Accuracy is crucial in research modeling the different outcomes of the observations, such as predicting which patients live *vs*. die. The second answer to the question is CTA models may be configured to maximize overall PAC, but research in this area has only recently begun[14,15] so little is yet known. A third reason is that relatively few phenomena follow a *linear* path. Rather, even simple real-world phenomena often trace a *lattice* (classification tree) or *circuitous* path to their final disposition. A fourth explanation is that MultiODA requires greater computing power *vs*. CTA. And, a fifth response is that CTA *avoids paradoxical confounding* by relaxing the legacy assumption that requires *exactly two* homogeneous groups exist in the optimal solution.[16]

Nevertheless, in addition to the earlier discussion, significant reasons exist to further study the MultiODA algorithm. For example, some class variables may be (nearly) perfectly modeled for a subset of the sample by using MultiODA with two attributes. This model-based variable may be represented *as a single attribute* (i.e., a "composite variable") within a node of an ODA or CTA model. Two-attribute MultiODA models may be obtained for much larger samples than models with more attributes. *If attributes are unit-weighted* then MultiODA problems involving a million observations and five attributes can be solved in CPU *seconds* by a mainframe computer.[3] An accurate MultiODA endpoint with a large N, stable in LOO analysis, may similarly be treated as a single attribute in structural decomposition analysis.[17]

Thus, code that can be used to conduct MultiODA analyses using modern computer systems is clearly needed.

References

[1]Soltysik RC, Yarnold PR (1994). The Warmack-Gonzalez algorithm for linear two-category multivariable optimal discriminant analysis. *Computers and Operations Research*, *21*, 735-745.

[2]Soltysik RC, Yarnold PR (2010). Two-group MultiODA: Mixed-integer linear programming solution with bounded *M*. *Optimal Data Analysis*, *1*, 31-37.

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

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

[5]Yarnold PR, Soltysik RC, Martin GJ (1994). Heart rate variability and susceptibility for sudden cardiac death: An example of multivariable optimal discriminant analysis. *Statistics in Medicine*, *13*, 1015-1021.

[6]Yarnold PR, Soltysik RC (2016). *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286

[7]Yarnold PR (1996). Discriminating geriatric and non-geriatric patients using functional status information: An example of classification tree analysis via UniODA. *Educational and Psychological Measurement*, *56*, 656-667.

[8]Yarnold PR, Bryant FB (2015). Obtaining a hierarchically optimal CTA model via UniODA software. *Optimal Data Analysis*, *4*, 36-53.

[9]Yarnold PR, Bryant FB (2015). Obtaining an enumerated CTA model via automated CTA software. *Optimal Data Analysis*, *4*, 54-60.

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

[11]Yarnold PR, Linden A (2016). Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis, *Optimal Data Analysis*, *5*, 65-73.

[12]Yarnold PR (2018). Visualizing application and summarizing accuracy of ODA models. *Optimal Data Analysis*, *7*, 85-89.

[13]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books.

[14]Yarnold PR (2016). Identifying the descendant family of HO-CTA models by using the minimum denominator selection algorithm: Maximizing ESS versus PAC. *Optimal Data Analysis*, *5*, 53-57.

[15]Yarnold PR (2016). Pruning CTA models to maximize PAC. *Optimal Data Analysis*, *5*, 58-61.

[16]Yarnold PR (2020). What is novometric data analysis? *Optimal Data Analysis*, *9*, 195-206.

[17]Yarnold PR (2015). UniODA-based structural decomposition *vs*. legacy linear models: Statics and dynamics of intergenerational occupational mobility. *Optimal Data Analysis*, *4*, 194-196.

Author Notes

No conflicts of interest were reported.