

Abstract

Dusseldorp, E. & Meulman, J. J. (2001). Prediction in medicine by integrating regression trees into regression analysis with optimal scaling. *Methods of Information in Medicine*, 40, 403-409.

Objectives: A new data-analysis strategy is proposed to solve the problems of selecting interaction terms in linear regression on the one hand, and of statistically testing the significance of regression trees on the other hand.

Method: The proposed strategy combines two data mining techniques: regression trees and regression analysis with optimal scaling (CATREG). The method traces small regression trees using the bootstrap and integrates the results as interaction variables (called “trunk variables”) into CATREG.

Results: An application to data from cardiac patients shows a relative increase of 19% variance accounted for (16% cross-validated variance), by the CATREG model including the trunk variables compared to the model excluding these variables.

Conclusion: This study indicates that trunk variables can be useful to model interaction effects in prediction problems.