Accepted Manuscript

Deriving optimal data-analytic regimes from benchmarking studies

Lisa L. Doove, Tom F. Wilderjans, Antonio Calcagnì, Iven Van Mechelen

PII: S0167-9473(16)30243-2
DOI: http://dx.doi.org/10.1016/j.csda.2016.10.016
Reference: COMSTA 6366

To appear in: *Computational Statistics and Data Analysis*

Received date: 1 December 2015
Revised date: 16 October 2016
Accepted date: 17 October 2016

Please cite this article as: Doove, L.L., Wilderjans, T.F., Calcagnì, A., Van Mechelen, I., Deriving optimal data-analytic regimes from benchmarking studies. *Computational Statistics and Data Analysis* (2016), http://dx.doi.org/10.1016/j.csda.2016.10.016

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

*Manuscript
Click here to view linked References

Deriving Optimal Data-Analytic Regimes from Benchmarking Studies

Lisa L. Doove[a,*], Tom F. Wilderjans[a,b], Antonio Calcagnì[a,c], Iven Van Mechelen[a]

[a]*Department of Psychology, Katholieke Universiteit Leuven, Tiensestraat 102 - bus 3713, Leuven, Belgium*
[b]*Methods and Statistics Unit, Institute of Psychology, Faculty of Social and Behavioral Sciences, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, the Netherlands*
[c]*Department of Psychology and Cognitive Science, University of Trento, Corso Bettini, 31 - 38068 Rovereto, Italy*

Abstract

In benchmarking studies with simulated data sets in which two or more statistical methods are compared, over and above the search of a universally winning method, one may investigate how the winning method may vary over patterns of characteristics of the data or the data-generating mechanism. Interestingly, this problem bears strong formal similarities to the problem of looking for optimal treatment regimes in biostatistics when two or more treatment alternatives are available for the same medical problem or disease. It is outlined how optimal data-analytic regimes, that is to say, rules for optimally calling in statistical methods, can be derived from benchmarking studies with simulated data by means of supervised classification methods (e.g., classification trees). The approach is illustrated by means of analyses of data from a benchmarking study to compare two different algorithms for the estimation of a two-mode additive clustering model.

Keywords: Additive clustering, Benchmarking, Classification trees, Comparison of methods, Data-analytic regimes, Supervised learning

*Corresponding author: Phone: +32 16 3 25977
Email addresses: lisa.doove@kuleuven.be (Lisa L. Doove), t.f.wilderjans@fsw.leidenuniv.nl (Tom F. Wilderjans), antonio.calcagni@unitn.it (Antonio Calcagnì), iven.vanmechelen@kuleuven.be (Iven Van Mechelen)

Download English Version:

https://daneshyari.com/en/article/4949366

Download Persian Version:

https://daneshyari.com/article/4949366

Daneshyari.com