Accepted Manuscript

Model-based clustering for assessing the prognostic value of imaging biomarkers and mixed type tests

Zheyu Wang, Krisztian Sebestyen, Sarah Monsell

Please cite this article as: Wang, Z., Sebestyen, K., Monsell, S., Model-based clustering for assessing the prognostic value of imaging biomarkers and mixed type tests. *Computational Statistics and Data Analysis* (2016), http://dx.doi.org/10.1016/j.csda.2016.10.026

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.

Model-based clustering for assessing the prognostic value of imaging biomarkers and mixed type tests

Zheyu Wang[a,*], Krisztian Sebestyen[a] and Sarah Monsell[b]

[a]Johns Hopkins University, Baltimore, Maryland, U.S.A.
[b]University of Washington, Seattle, Washington, U.S.A.
*email: wangzy@jhu.edu

Abstract

A model-based clustering method is proposed to address two research aims in Alzheimers disease (AD): to evaluate the accuracy of imaging biomarkers in AD prognosis, and to integrate biomarker information and standard clinical test results into the diagnoses. One challenge in such biomarker studies is that it is often desired or necessary to conduct the evaluation without relying on clinical diagnoses or some other standard references. This is because 1) biomarkers may provide prognostic information long before any standard reference can be acquired; 2) these references are often based on or provide unfair advantage to standard tests. Therefore, they can mask the prognostic value of a useful biomarker, especially when the biomarker is much more accurate than the standard tests. In addition, the biomarkers and existing tests may be of mixed type and vastly different distributions. A model-based clustering method based on finite mixture modeling framework is introduced. The model allows for the inclusion of mixed typed manifest variables with possible differential covariates to evaluate the prognostic value of biomarkers in addition to standard tests without relying on potentially inaccurate reference diagnoses. Maximum likelihood parameter estimation is carried out via the EM algorithm. Accuracy measures and the ROC curves of the biomarkers are derived subsequently. Finally, the method is illustrated with a real example in AD.

Keywords: finite mixture, latent variable model, diagnostic tests, biomarkers, differential covariate effect, imperfect gold standard