Editorial

Hans-Hermann Bock

Published online: 12 January 2012
© Springer-Verlag 2012

This first issue of volume 6 (2012) of the journal *Advances in Data Analysis and Classification (ADAC)* comprises four articles which deal with clustering for mixed data types, the visual exploration of incomplete data, the analysis of multiset data by an extension of the classical STATIS-ACT method, and properties of Cohen's kappa for measuring agreement among raters. Proposed methods are typically illustrated by real-case or simulated examples.

In a first paper on '*A latent variables approach for clustering mixed binary and continuous variables with Gaussian mixture model*' Isabella Morlini addresses the classical problem of clustering mixed data types. Many methods have been proposed for this purpose, classical solutions comprise the transformation of all variables into the same type, typically a categorical or ordinal one. Morlini's approach proceeds in the inverse way insofar as it assumes that the discrete (binary) variables are generated from (latent) continuous variables by discretization with suitable thresholds, and the joint distribution of all (observed and latent) continuous variables is multivariate Gaussian. Morlini's method derives estimates for the characteristics of this Gaussian distribution (e.g., by involving the tetrachoric correlation coefficients), for the unknown scores of the latent variables, and for the thresholds for discretization. From these bases a mixture-type clustering method is designed. Its performance is illustrated by simulated and real-case examples and compared to other methods, e.g. with the classical 'Latent Gold package' and a mixture model with class-specific joint multivariate and multinomial distributions.

The second article, written jointly by Matthias Templ, Andreas Alfons and Peter Filzmoser, is devoted to the analysis of data sets with missing values, specifically to the problem of identifying possible structures of missing data. Under the title '*Exploring incomplete data using visualization techniques*' they present a range of visual methods for investigating the distribution of missing values and their relation to the observed data. The paper stresses the usefulness of visualization techniques. The authors show how a combined approach comprising visualization, imputation, and data analysis can be performed by using the R package VIM, without additional software.

The article ‘*Analyzing multiset data by the Power STATIS-ACT method*’ by Jacques Benasseni and Mohammed Bennani Dosse addresses the problem of simultaneously analyzing several sets of data, e.g., by Generalized Procrustes Analysis, generalized canonical analysis etc. STATIS-ACT is another strategy in this context that, given K data matrices of the same size, looks for a compromise matrix by minimizing a quadratic (‘power 2’) performance criterion. The proposed Power STATIS-ACT method considers a modified criterion where a general power parameter $s \geq 1$ is used (where $s = 1$ plays a special role). Theoretical and practical properties of this method are investigated, the results are illustrated by a number of real-case data tables.

This issue closes with a paper by Matthijs J. Warrens that investigates the structure of an $n \times n$ agreement table where two persons rate the same set of objects according to n ordered categories and the table displays the number or frequencies of pairwise ratings. In 1968 Cohen has proposed his well-known kappa coefficient for measuring the agreement of the raters on the basis of this table. Warrens considers the situation where adjacent categories are combined into m classes such that we obtain collapsed agreement tables of a lower size $m \times m$. Under the title ‘*Cohen’s linearly weighted kappa is a weighted average*’ he essentially shows that Cohen’s kappa can be obtained by averaging agreement measures related to these smaller $m \times m$ tables. Various versions of this result are formulated and proved. A range of numerical examples illustrates these theorems.

I conclude with a remark on the journal’s management from 2012. In fact, the journal ADAC has now existed for five years, with increasing size (from three to four issues per year) and considerable interest from the statistical community as is illustrated by the increasing number of downloads and paper submissions. This also means that, in accordance with the statutes of the ADAC Editorial Board, the initial structure of the board should be modified, updated, and enlarged in order to cope with the increasing workload.

Within this updating process I will hand over my role as the Coordinating Editor of ADAC to my colleague, Maurizio Vichi, who explains, in a separate note, various changes to be made in the future editorial structure, as well as in the paper submission procedure of ADAC. Given that this is the last issue of ADAC which is prepared under my responsibility I would like to thank my colleagues from the Editorial Board, Wolfgang Gaul, Maurizio Vichi, and Akinori Okada, for their continuous work, support, cooperation and enthusiasm in the process of initiating and running this journal, in particular by organizing reviewing processes for many manuscript submissions. Thanks also to the members of the Associate Editorial Board for their very much appreciated advice and suggestions. Last but not least I am very much indebted to dozens of anonymous reviewers of manuscripts and would like to sincerely acknowledge their time-consuming work and commitment.

I conclude with a heartfelt thank-you for Maurizio Vichi for his willingness to take over my coordinating activities from 2012 and look forward to a most challenging development of ADAC to the benefit of its authors and its readers.