Editorial

Published online: 6 August 2011
© Springer-Verlag 2011

The current issue 3 of volume 5 (2011) of the journal *Advances in Data Analysis and Classification (ADAC)* comprises four articles which deal with topics related to missing values in principal component analysis, consensus methods for partitioning, dimension reduction in classification, and rescaling of similarity measures. Proposed methods are typically illustrated by real-case or simulated examples.

In a first paper on ‘Correcting Jaccard and other similarity indices for chance agreement in cluster analysis’ *Ahmed N. Albatineh* and *Magdalena Niewiadomska-Bugaj* derive, for various classical similarity measures (e.g., the Jaccard, Tanimoto, Gower/Legendre, and Sokal/Sneath indices), rescaled versions that are ‘corrected’ for chance variations. This is realized by subtracting the expectation and dividing by the standard deviation under suitable probability models. The essential contribution of the paper consists in the computation or approximation of these distribution parameters and a detailed study on the performance of the corrected indices for simulated as well as for real-case data.

In the second article on ‘Linear dimension reduction in classification: adaptive procedure for optimum results’ *Karsten Luebke* and *Claus Weihs* start from the remark that there is a variety of techniques for reducing the dimension of a recorded data set before applying a classification rule. On the other hand, the performance of each technique depends on data characteristics that are a priori unknown for most data situations. So no ‘best’ dimension reduction method is known a priori. The authors propose, for finding a good compromise strategy, an adaptive (meta-learning) procedure that constructs, in a first step, rules based on simulated data which attempt to cover the space of realistic data configurations and, in a second step, to select the actual classification and dimension reduction method to be applied to the observed data, on the basis of these data and a so-called ‘selection statistic’. The performance of this two-step strategy is illustrated for artificial three- and four-class examples as well as for a real life macroeconomic data set dealing with German business cycles.

The article ‘Consensus of partitions: a constructive approach’ by *Alain Guénoche* is devoted to the problem of constructing, for a given family of partitions of a set of

items, a ‘consensus partition’ that contains a maximum number of joined or separated pairs of items that are also joined or separated in the family members. This problem is usually formulated as a maximization problem for a score function that can be resolved, in certain cases, by integer linear programming. In contrast, Guénoche’s paper proposes a polynomial heuristic maximization algorithm that can also be applied to large sets of items. It appears that the resulting pseudo-optimum partition is close (or even identical) to the optimum (consensus) partition in cases where the latter one can be explicitly computed.

The article by *Julie Josse, Jérôme Pagès* and *François Husson* on ‘Multiple imputation in principal component analysis’ considers the fact that the available methods to handle missing values in PCA do not take into account, and estimate, the additional variability of the results (point estimates of axes and components, estimates of missing values) that is implied by the occurrence of missing values. In order to tackle with this problem they propose to use multiple imputation methods. Two different strategies are explained, and the methods are illustrated and evaluated by simulated data.

Finally this issue contains a Call for Papers for a Special Issue of ADAC that is devoted to the domain *Data Analysis and Classification in Marketing*, scheduled for 2012. Guest editors will be Prof. Dr. Daniel Baier (Cottbus) and Prof. Dr. Reinhold Decker (Bielefeld). Researchers and practitioners who are working in this field are kindly invited to submit manuscripts for this Special Issue before the deadline December 31, 2011.

Hans-Hermann Bock (Aachen)
Wolfgang Gaul (Karlsruhe)
Akinori Okada (Tokyo)
Maurizio Vichi (Rome)