Editorial

Published online: 3 May 2012
© Springer-Verlag 2012

This second issue of volume 6 (2012) of the journal *Advances in Data Analysis and Classification* (ADAC) comprises four articles which deal with dissimilarity and similarity measures for comparing dendrograms, robust clustering, orthogonal rotation in a principal component method for a mixture of qualitative and quantitative variables, and the symbolic analysis of histogram data by means of a modified interval principal component analysis (PCA). Proposed methods are usually illustrated by real-case or simulated examples.

In the first paper on ‘*Dissimilarity and similarity measures for comparing dendrograms and their applications*’ Isabella Morlini and Sergio Zani propose a new measure to address the problem of comparing and assessing differences in two hierarchical classifications (dendrograms) of n objects. Some properties of this new measure are: it is a metric measure, normalized in the interval $[0, 1]$, it can be decomposed into the $(n - 2)$ dissimilarities between nested partitions forming the two compared dendrograms of n objects, it does not depend on the group labels since it refers to pairs of objects only. When the new distance between dendrograms is compared with dissimilarity measures previously proposed in the literature, the new distance has at least three major advantages. First it is a real-valued metric and this is useful in the context of consensus of classifications. Second it can be used to measures the distance of any two sequences of partitions of n objects into $2, 3, \dots, n - 2, n - 1$ groups (so the fact that the partitions originate from successively cutting two dendrograms is irrelevant). The partitions could also be non-nested ones. Thus, for the new distance the contribution of each partition in determining differences between two dendrograms can be assessed. Third, the new distance can also be applied to compare two sequences of clusterings with different numbers of groups, for example corresponding to cutting one tree at 2, 3, 4 clusters and the other tree at 3, 4, 5 clusters.

The second article, on “*The influence function of the TCLUS robust clustering procedure*”, written jointly by Christel Ruwet, Luis Angel García-Escudero, Alfonso

Gordaliza, and Agustin Mayo-Isar, deals with the problem to study robustness properties of the TCLUST procedure by means of the influence function. TCLUST allows to perform heterogeneous robust clustering, i.e. Cluster Analysis with groups having different scatter structures and weights and allowing also the presence of outliers.

Several interesting results are discussed in this paper. The influence functions of the TCLUST statistical functional is bounded, thus, infinitesimal contamination concentrated at a given point cannot have a devastating impact on the estimation of the TCLUST parameters. Furthermore, contamination outside clusters reduces the difference between weights, moves the centers away from each other and increases scatters of clusters. The contamination in a cluster increases the weight of this cluster while it decreases the weight of other ones and it has a bigger impact on the center of this cluster than on the other centers. More generally, contamination has a bigger impact on the estimation in the smallest cluster in cases of separated and overlapping clusters or in the widest cluster in the case of inside clusters. The contamination in a cluster reduces the scatter of the other clusters except if there is overlap between clusters. The contamination near the center of a cluster decreases the scatter of this cluster. Finally, when the Eigenvalues Ratio restriction (ER) is enforced, the influence functions of the scatter statistical functionals are linked, resulting in a behavior at the clusters boundaries different from that without the restriction.

In the third paper on “*Orthogonal rotation in PCAMIX*”, Marie Chavent, Vanessa Kuentz-Simonet, and Jérôme Saracco first suggest a new original presentation of PCAMIX, a well-known principal component method for a mixture of qualitative and quantitative variables, where the principal components and the squared loadings are obtained from a Singular Value Decomposition. PCAMIX includes the ordinary PCA and multiple correspondence analysis (MCA) as special cases. This new formulation of PCAMIX leads to a computationally efficient procedure for varimax rotation in PCAMIX and to a direct solution for the optimal angle of rotation. A simulation study shows the good computational behavior of the new algorithm with respect to the previously proposed procedure for varimax rotation for PCAMIX.

The fourth paper on “*Adaptation of interval PCA to symbolic histogram variables*” written by Sun Makosso-Kallyth and Edwin Diday, deals with an adaptation of symbolic interval PCA to histogram data. Two methodologies have been proposed. The first one includes three steps: the coding of bins of histogram, the ordinary PCA of means of variables and the graphical representation of dispersion of symbolic observations. For this latter one a transformation of histograms into intervals is proposed. Then, the projection of the hypercubes or the lengths of intervals associated to each symbolic observation on the principal axes of the ordinary PCA of means is suggested. In the second methodology, the use of the three previous steps with the angular transformation is suggested. The proposed methodology is compared with factorial multiple analysis and the STATIS method by using the well-known Iris Data set.

Finally, this issue contains the Call for Papers for a Special Issue of ADAC that is devoted to the domain *Model Based Clustering and Classification* and scheduled for the end of 2013. Guest editors will be Hans-Hermann Bock (Aachen, Germany),

Salvatore Ingrassia (Catania, Italy), and Jeroen Vermunt (Tilburg, The Netherlands). Researchers and practitioners who are working in this field are kindly invited to submit original manuscripts for this Special Issue before the deadline November 30, 2012.

Hans-Hermann Bock (Aachen)

Wolfgang Gaul (Karlsruhe)

Akinori Okada (Tokyo)

Maurizio Vichi (Rome)

Claus Weihs (Dortmund)