Journal of Classification 31:135 (2014)
DOI: 10.1007/s00357-014-9160-0

Editorial

The first article in this issue, by Andrews and McNicholas, discusses a relatively simple and general variable selection algorithm suitable for both supervised and unsupervised classification, which strikes a good balance between minimizing within-group variance and between-variable correlation. The authors show that it is competitive with some of the existent variable selection techniques.

In the second paper, on time series classification, Bagnall and Janacek propose a run-length transformation that can detect similarity in auto-correlation between two series without having to form the complete auto-correlation function. They show that it can successfully distinguish data from different stationary auto-regressive models. In the third paper, Warrens establishes the relation between several measures of the amount of agreement among independent ordered classifications, and shows that the widely used weighted kappa methodology could perhaps better be replaced with the Zegers and Ten Berge family of coefficients, corrected for chance agreement, which includes the standard product-moment correlation.

The next two articles consider two major clustering situations. First, Lee and Wilcox discuss generalizations of the classic Ward agglomerative hierarchical clustering method, in which powers of Minkowski distances are used instead of squared Euclidean distances, and show that they keep the desirable properties of the Lance-Williams updating formula, while often giving better clustering results. Especially the simple Manhattan metric appears to perform well in the Ward algorithm. Next, Carbonneau, Caporossi and Hansen look at cluster-wise regression, in which different regression lines or surfaces are fitted for several subgroups of observations that are unknown in advance. They propose a special branch and bound strategy that finds globally optimal solutions and is competitive with existing schemes as the number of clusters increases beyond three.

Van Ginkel and Kroonenberg show in the last article of this issue how to do multiple imputation of missing data in the context of principal components analysis, where the problem of combining the different results obtained under different imputations is solved by generalized Procrustes analysis. This strategy resolves the problems encountered in earlier procedures for the same purpose, such as the mean Varimax method.

The issue concludes with a book review by Ramsés Mena of Kadane's recent text on Bayesian statistics.

Willem J. Heiser