Contributions of Italian Statisticians to the Development of Multivariate Data Analysis

Renato Coppi and Paolo Giordani

Abstract

The main contributions of Italian statisticians to the methodology of multivariate data analysis are investigated, focusing specifically on the development of techniques for coping with the extraction of information from complex data characterized by two or more variables or sets of variables as observed on one or more sets of objects. In particular the following types of methodological areas are considered: supervised and unsupervised classification, regression, factorial and scaling approaches. Although the bulk of this study is devoted to the works appeared in the last three or four decades, some hints are given to the historical profile of the Italian school of Statistics. In this connection it is underlined that the more recent developments are characterized by specific traits of originality, which place the Italian contributions to the aforementioned fields of research somehow at the crossroads among the French, the Dutch, and the Anglo-American schools of Statistics.

1 Introduction

Data analysis has historically been characterized by a "dualistic" perspective. On one side, statistical data have been looked on as being generated by a probabilistic model most often expressed in parametric terms. Typically, inferential procedures based on the likelihood have been adopted for analyzing the empirical information conveyed by the data. The Anglo-American school of thought has greatly influenced this approach. On the other side, the data have been viewed as containing an intrinsic information, independently of any prior knowledge concerning their generation

R. Coppi (✉) • P. Giordani
Dipartimento di Scienze Statistiche, Sapienza Università di Roma, Rome, Italy
e-mail: renato.coppi@uniroma1.it; paolo.giordani@uniroma1.it

F. Crescenzi and S. Mignani (eds.), *Statistical Methods and Applications from a Historical Perspective*, Studies in Theoretical and Applied Statistics,
DOI 10.1007/978-3-319-05552-7__2,

process. In this case, the analytical procedures have been devised in such a way as to discover and display the empirical information by means of appropriate representations of the data in geometric or algebraic structures, utilizing the formal properties of these structures for carrying out the analyses [Principal Component Analysis (PCA), Correspondence Analysis, Cluster Analysis are examples of this approach, if we look at them from an "exploratory" viewpoint]. The French school of "Analyse des Données" has typically represented this line of research particularly in the period 1960–1980. A third perspective has appeared in the literature in the 1980s and 1990s, mainly inspired by the Dutch school of data analysis (Gifi 1990). This may be defined as "anti-dogmatic," since it is based on flexibility and eclecticism. Although the observed data must drive the procedures of information extraction, an instrumental role is also assigned to statistical models including probabilistic mechanisms that enable the investigator to validate and generalize the results obtained from the analysis of the available data. Alternating least squares algorithms constitute the basic computational tools for implementing the analytical procedures; resampling techniques are often adopted for testing and validation. The school of "Statistical Learning" (Hastie et al. 2009; Vapnik 1986) has many points in common with the latter line of thought, although it puts more emphasis on the probabilistic tools due to the fact that an underlying unknown stochastic process of data generation is assumed and the predictive perspective of the analysis is strongly underlined.

In our paper, we focus on the contributions of Italian statisticians to the domain of multivariate data analysis, restricting our attention to the developments of the "exploratory" approach, including the possible use of probabilistic tools (an interesting previous review of the Italian contributions to this field of analysis is by Balbi 1994). In this respect, it can be stated that the lines of research of Italian scholars have taken inspiration from each of the abovementioned schools of thought, thus realizing a sort of compromise among different ways of looking at the data, with the aim of enhancing and representing in various manners the information embodied in the empirical observations. In this connection, it is evident the link with the traditional characteristics of the Italian school of Descriptive Statistics. One basic feature of this school has historically been the endeavor to incorporate in the statistical tools for analyzing real-world phenomena their *complexity*, without introducing too many theoretical assumptions. While in the first half of the twentieth century, this attitude has mainly produced a wide range of descriptive statistical indices for both univariate and multivariate set-ups, in the second part of the century and more remarkably in the last three decades it turned towards the construction of methods for handling the information contained in complex dependence and interdependence structures among and between sets of variables and for detecting meaningful typologies of statistical units. At the same time, a specific interest has been focused on the analysis of what can be called "complex" statistical observations. In this category we include multiway data arrays as well as special empirical objects like interval-valued or fuzzy-valued variables or, more generally, symbolic data.

In the sequel we will mention some of the numerous contributions provided by Italian statisticians to the above fields of methodological research. In doing that we will limit ourselves to the works whose inspiration is more close to the exploratory perspective rather than to the inferential probabilistic viewpoint. Of course this does not exclude that some of the considered contributions are, in some sense, "model-based" and make use of probabilistic tools. As a matter of fact, this use might be interpreted in the abovementioned framework of the Dutch school, i.e. just as technical means for drawing information from the data and possibly evaluating its statistical reliability. Another preliminary remark concerns the selection of works that we will consider. Due to the practical impossibility of covering the great deal of interesting contributions in the fields under investigation, we have limited the illustration to a small part of the scientific production of interest, trying to enhance the main lines of research rather than detailing the specific proposals. Therefore, we apologize in advance for all the citations that have been missed, while we recognize the value of the overall contribution of Italian scholars to the domain of multivariate data analysis, with particular reference to the many statisticians involved in the activity of the CLADAG section of the Italian Statistical Society. The following discussion is divided into four parts, referring, respectively, to: Regression and Classification, Association structures, Interval and Fuzzy Data, Multiway Data. Due to the limited size of the present paper, the illustration will be very schematic, thus the reader is invited to look at the bibliographic references for getting a deeper insight into the various topics.

2 Regression and Classification

Several lines of research have been followed in this domain, with the common aim of taking into account the complexity of real-world dependence structures for both quantitative and qualitative response variables. In this perspective, some of the works look for improving methods and techniques of common use, while some other works suggest new approaches or propose new methods of analysis.

Noticeable contributions to classification and regression procedures based on decision trees have been provided by Conversano, Cappelli, Mola, and Siciliano (see, e.g., Cappelli et al. 2002; Conversano and Siciliano 2009; Siciliano and Mola 2000). These contributions concern several aspects, including the improvements of decision trees methodology and its use in various data analysis situations (data mining, imputation of missing data, etc.). In this area several other authors provided interesting proposals. For instance, Miglio and Soffritti (2005) introduced suitable proximity measures between Classification trees, which can also be utilized for finding parsimonious optimal trees from among a set of possible trees (Miglio and Soffritti 2004). Another direction of research consists in enlarging the scope of the classical CART trees to include, for example, ordinal response variables. Suitable splitting criteria have been introduced by Piccarreta (2004) to this aim, utilizing a new measure of association between categorical and ordinal variables (Piccarreta 2001).

Concerning traditional algorithms of cluster analysis, such as the k-means method, various authors have provided useful suggestions for improving their efficiency in detecting underlying typologies which can be captured by looking at the empirical features shown by the distribution of the observed units in the space. The use of appropriate metrics and of suitable weights for the observations allows for remarkable improvements in this direction (see, e.g., Cerioli 2001; Cerioli and Zani 2001). The problem of a simultaneous dimensionality reduction and clustering procedure has been faced by Vichi, Rocci, and Vicari in several works, with reference to units $\times$ variables matrices (see, e.g., Rocci et al. 2011; Timmerman et al. 2010; Vichi and Kiers 2001) and also with respect to proximity matrices by means of an appropriate utilization of multidimensional scaling (Kiers et al. 2005). The problem of clustering variables rather than units has been, for instance, faced by Laghi and Soffritti (2005) who proposed a procedure based on suitable measures of collinearity within groups (clusters). Along this line of research a generalized double k-means technique has also been proposed by Vichi and Rocci for simultaneously clustering the units on one side and the variables on the other side (Vichi and Rocci 2008; Vichi and Saporta 2009). In the field of model-based clustering, Galimberti and Soffritti (2007) introduced model-based procedures for detecting multiple cluster structures, namely typologies of units based on different subsets of a set of observed variables.

A new perspective for regression and classification studies was introduced by the Statistical Learning approach, which puts particular emphasis on the predictive viewpoint and on the use of computer intensive techniques based on boosting, bagging and random forests procedures, enabling the researcher to improve the predictive efficiency of regression and classification models. Systematic contributions to this line of research have been provided by Di Ciaccio and Borra in several works. For instance, in Borra and Di Ciaccio (2002) they use bagging and boosting for improving the prediction capability of non-parametric regression techniques. In Borra and Di Ciaccio (2010) they investigate various ways of measuring the prediction error in regression and classification problems and propose some methodological improvements. In the same line, other authors, like Sandri and Zuccolotto (2006), suggest how to use random forests for selecting predictors in a classification problem. Giordano et al. (2004) devise suitable resampling procedures for selecting variables in neural network models for regression analysis and, more generally, contribute to the regression and classification methods based on neural networks (e.g., Perna and Giordano 2001).

3 Association Structures

This is a wide field of analysis whose objective is to detect and display observable or latent structures of association within and between sets of variables. In the following we will just mention some interesting lines of research to which the Italian statisticians gave a noticeable contribution.

Starting from the seminal paper by D'Ambra and Lauro (1982) concerning the analysis of multivariate data with respect to reference subspaces, a line of research has been developed in the field of association structures, consisting in an asymmetric approach to the analysis of relationships among several variables taking into account the influence of "external" variables. The paper by Lauro and D'Ambra (1992) well illustrates this line of thought, while in the works by Beh and D'Ambra (2009); Lombardo et al. (2007, 1996) it is shown how the application of this approach to three-way contingency tables gives rise to a new method called nonsymmetric correspondence analysis, allowing to deal with ordinal variables. In the last decades, different authors tried to decompose association measures for three-way contingency tables. In this perspective, papers by Italian scholars involved in the analysis of three-way contingency tables can be found in Beh et al. (2007); Lombardo (2011); Lombardo et al. (1996); Siciliano and Mooijaart (1997); Simonetti et al. (2010).

Another line of research refers to the improvement and utilization of the Partial Least Squares (PLS) approach in the analysis of several types of association structures. Esposito Vinzi and various co-authors introduce the "Generalized PLS Regression" model (Bastien et al. 2005), which basically allows a transformation of the explanatory variables into orthogonal components in order to improve the fitting to a quantitative or qualitative response variable, by using an appropriate iterative non-linear least squares algorithm. Esposito Vinzi, Lauro and others (Tenenhaus et al. 2005) discuss and improve the procedures of "PLS path modelling," which adopts the PLS approach for estimating suitable structural equation models. Moreover, the authors show that it can provide a general framework for analyzing a multi-block structure of observed variables. A special issue of Computational Statistics and Data Analysis (Esposito Vinzi and Lauro 2005), edited by Esposito Vinzi and Lauro, has been devoted to the various methodological improvements of the PLS approach in several areas of multivariate analysis.

Multidimensional Scaling has been also studied by Italian statisticians. Combining the information contained in the proximities between statistical units with the values taken by "external" variables on those units may lead to a deeper analysis of association through the use of multidimensional scaling procedures which allow also useful visualizations of the results. This line of research is witnessed, for instance, by the works of Bove and Rocci with particular reference to asymmetric proximities (see, e.g., Rocci and Bove 2002).

4 Interval and Fuzzy Data

Interval data, fuzzy data and, more generally, symbolic data represent complex observations requiring particular techniques of analysis. This "complexity" may be due to: (1) imprecision/vagueness of the observed variables, (2) intrinsic complexity of the observed phenomena (represented as "symbolic objects"). Source (1) constitutes a type of uncertainty which needs special methods of treatment, (2) involves a specific mathematical representation of "non-standard" variables. We focus our

illustration on type (1) data, which represent anyway relevant instances of symbolic data (to whose study are addressed many works, in particular, by Verde; see, e.g., Irpino and Verde 2008; Lauro et al. 1998).

Italian statisticians have given remarkable contributions to the analysis of interval data, generalizing in various ways the techniques of PCA, Cluster Analysis and Regression, in order to make them suitable for the analysis of this type of complex data. Concerning PCA, Lauro and Palumbo (2000, 2003) and subsequently Gioia and Lauro (2005) have worked on intervals represented as boxes (or hyperrectangles) in Euclidean spaces looking at their vertices and ranges. Coppi et al. (2003) and D'Urso and Giordani (2004) utilize a "data reconstruction" approach based on a Midpoint-Radius representation of the intervals. Approaches to cluster analysis of interval data have been proposed by, for instance, D'Urso and Giordani (2006) and Irpino and Tontodonato (2006).

Also in the field of fuzzy data many important contributions of the Italian statisticians are to be recorded. A systematic program of re-interpretation of the classical PCA, clustering and regression techniques in terms of fuzzy data is being realized. In particular, Coppi, D'Urso, Ferraro, and Giordani have produced a series of papers in this direction. Basic ingredients of these extensions are: (1) the formalization of fuzzy observations in terms of LR fuzzy numbers; (2) the introduction of appropriate metrics for LR fuzzy variables; (3) the construction of specific models for each of the above fields of analysis (PCA, cluster analysis, regression), incorporating the fuzziness of the observations; (4) the definition of suitable algorithms for estimating the parameters of the introduced models (generally iterative least squares procedures); (5) the possible utilization (in particular for regression analysis) of the notion of Fuzzy Random Variable, which enables the investigator to cope simultaneously with the uncertainty stemming from the imprecision of the data, and the one due to an assumed probabilistic mechanism generating the observations. A partial list of the contributions provided in the above framework is as follows. Concerning PCA of fuzzy data (Coppi et al. 2006b; Giordani and Kiers 2004). For cluster analysis of fuzzy data and fuzzy multivariate trajectories (Coppi et al. 2012; D'Urso and Giordani 2006). As to regression analysis with fuzzy response and crisp or fuzzy explanatory variables (Coppi et al. 2006; D'Urso 2003; Ferraro et al. 2010). Finally, it must be underlined the publication of a special issue of Computational Statistics and Data Analysis, edited in 2006 by Coppi et al. (2006a), devoted to the recent developments of Fuzzy Statistical Analysis.

5 Multiway Analysis

Data generally refer to the observations of some variables on a set of units and are stored in a (two-way) matrix. However, in several situations, data on a set of units on some variables are assumed to be collected in different occasions, leading to a (three-way) array.

In this section, we shall limit our attention to the Italian contributions to the topic of multiway data analysis distinguishing two lines of research, namely component models and cluster analysis.

With respect to component models, at the beginning, the Italian statisticians involved in multiway analysis took inspiration from the French school. In this connection we cite the works by Bolasco (1986), Coppi (1986), and Coppi and Zannella (1978). Notice that in Coppi and Zannella (1978) the so-called Dynamic Factor Analysis has been introduced. Later, it has been extended and generalized (see Coppi and D'Urso 2002; Coppi et al. 1999; Corazziari 1999). In the following years, the Italian statisticians spread their research interests to the methods inspired also from the other schools acquiring an impressive importance within the world multiway community. The success of the meeting "Multiway'88" held in Rome bears witness to the Italian contributions to multiway analysis. The proceedings of the conference (Coppi and Bolasco 1989) represent a milestone within the domain and witness the active role played by Italian researchers. The most relevant findings of Italian statisticians to the Tucker3 and Candecomp/Parafac models and related methods were given by Rocci (1992), Rocci and Giordani (2010), Rocci and Ten Berge (2002). Extensions of such models to imprecise and/or vague data were proposed in Giordani (2010), Giordani and Kiers (2004). Other contributions on component methods can be found in Rizzi and Vichi (1995a,b). Finally, it is useful to mention the special issue of Computational Statistics and Data Analysis on multiway models edited by Coppi and Di Ciaccio (1994), which offered the current (in 1994) world state-of-art of multiway analysis.

The clustering problem for multiway data has been deeply addressed by Italian statisticians. As far as we know, Rizzi (1989) was the first one involved. At least three lines of research can be highlighted. The first one concerns the attempt to look for a sort of compromise partition built on the basis of the set of partitions for all the occasions. Therefore, the research interest is to find the single partition synthesizing at best, according to a given criterion, K available two-way partitions (one for each occasion). In this connection the contributions of Vichi deserve to be cited (Gordon and Vichi 2001; Vichi 1998). Another approach consists of seeking clusters of units considering the data array as a whole, in the sense that no distinct partitions for each occasion are assumed. Thus, in this respect, the data taxonomy is sought by considering the features of the units as the information on a number of variables in different occasions. Rocci, Vicari, and Vichi played an active role in this domain (Rocci and Vichi 2005; Vicari and Vichi 2009; Vichi et al. 2007). Notice that a few of these papers also involved a reduction of the entities of the variable and occasion modes through a Tucker3 model. Finally, the third line of research is about fuzzy clustering for time trajectories. In this case, the occasion mode is the time and data consist of the same variables observed on a number of units in different time occasions. Every unit can then be seen as a multivariate time trajectory. The aim of the analysis is to find a limited number of clusters composed by trajectories homogeneous according to suitable dissimilarity measures. Findings in this domain are due mainly to D'Urso (2000, 2004) also in collaboration with Coppi and D'Urso (2003), Coppi et al. (2010).

References

Balbi, S.: L'Analisi Multidimensional del Dati negli Anni Novanta. Rocco Curto Editore, Napoli (1994)

Bastien, P., Esposito Vinzi, V., Tenenhaus, M.: PLS generalized regression. Comput. Stat. Data Anal. **48**, 17–46 (2005)

Beh, E., D'Ambra, L.: Some interpretative tools for non-symmetrical correspondence analysis. J. Classif. **27**, 55–76 (2009)

Beh, E., Simonetti,B., D'Ambra, L.: Partitioning a non-symmetric measure of association for three-way contingency tables. J. Multivar. Anal. **98**, 1391–1411 (2007)

Bolasco, S.: Per una teoria sulla costruzione e l'analisi delle matrici a tre modi. In: Proceedings of the 33rd Meeting of the Italian Statistical Society, pp. 183–195 (1986)

Borra, S., Di Ciaccio, A.: Improving nonparametric regression methods by bagging and boosting. Comput. Stat. Data Anal. **38**, 407–420 (2002)

Borra, S., Di Ciaccio, A.: Measuring the prediction errors: a comparison of cross-validation, bootstrap and covariance penality methods. Comput. Stat. Data Anal. **54**, 2976–2989 (2010)

Cappelli, C., Mola, F., Siciliano, R.: A statistical approach to growing a reliable honest tree. Comput. Stat. Data Anal. **38**, 285–299 (2002)

Cerioli, A.: Elliptical clusters and the K-means algorithm. In: Book Short Papers CLADAG, Istituto di Statistica, Università degli Studi di Palermo, Palermo, pp. 13–16 (2001)

Cerioli, A., Zani, S.: Exploratory methods for detecting high denosty regions in clusters analysis. In: Borra, S., Rocci, R., Vichi, M., Schader, M. (eds.) Advances in Classification and Data Analysis, pp. 11–18. Springer, Berlin (2001)

Conversano, C., Siciliano, R.: Incremental tree-based missing data imputation with lexicographic ordering. J. Classif. **26**, 361–379 (2009)

Coppi, R.: Analysis of three-way data matrices based on pairwise relation measures. In: De Antoni, F., Lauro, C.N., Rizzi, A. (eds.) Proceedings in Computational Statistics, pp. 129–139. Physica-Verlag, Wien (1986)

Coppi, R., Bolasco, S. (eds.): Multiway Data Analysis. North Holland, Amsterdam (1989)

Coppi, R., Di Ciaccio, A. (eds.): Multiway data analysis: software and applications (Special issue). Comput. Stat. Data Anal. **18**(1), 3–184 (1994)

Coppi R., D'Urso, P.: The dual dynamic factor analysis model. In: Gaul, W., Ritter, G. (eds.) Classification, Automation, and New Media, pp. 47–55. Springer, Heidelberg (2002)

Coppi, R., D'Urso, P.: Three-way fuzzy clustering models for LR fuzzy time trajectories, Comput. Stat. Data Anal. **43**, 149–177 (2003)

Coppi, R., Zannella, F.: L'analisi fattoriale di una serie temporale multipla relativa allo stesso insieme di unità statistiche. In: Proceedings of the 29th Meeting of the Italian Statistical Society, pp. 61–77 (1978)

Coppi, R., Blanco, J., Camaño, G., Corazziari, I.: Descomposición factorial y regresiva de la variabilidad de un array a tres vias. Quantum **4**, 81–107 (1999)

Coppi, R., D'Urso, P., Giordani, P.: Data reduction models for interval valued observations. In: Book of Short Papers CLADAG 2003 (CLADAG'2003, Bologna), pp. 119–122 (2003)

Coppi, R., D'Urso, P., Giordani, P., Santoro, A.: Least squares estimation of a linear regression model with LR fuzzy response. Comput. Stat. Data Anal. **51**, 267–286 (2006)

Coppi, R., Gil, M.A., Kiers, H.A.L. (eds.): The fuzzy approach to statistical analysis (Special issue). Comput. Stat. Data Anal. **51**(1), 1–452 (2006)

Coppi, R., Giordani, P., D'Urso, P.: Component models for fuzzy data. Psychometrika **71**, 733–761 (2006)

Coppi, R., D'Urso, P., Giordani, P.: A fuzzy clustering model for multivariate spatial time series. J. Classif. **27**, 54–88 (2010)

Coppi, R., D'Urso, P., Giordani, P.: Fuzzy and possibilistic clustering for fuzzy data. Comput. Stat. Data Anal. **56**, 915–927 (2012)

Corazziari, I.: Dynamic Factor Analysis. In: Vichi, M., Opitz, O. (eds.) Classification and Data Analysis, Theory and Application, pp. 171–178. Springer, Heidelberg (1999)
D'Ambra, L., Lauro, N.C.: Analisi in componenti principali in rapporto ad un sottospazio di riferimento. Riv. Stat. Appl. **15**, 51–67 (1982)
D'Urso, P.: Dissimilarity measures for time trajectories. J. Ital. Stat. Soc. **1–3**, 53–83 (2000)
D'Urso, P.: Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data. Comput. Stat. Data Anal. **42**, 47–72 (2003)
D'Urso, P.: Fuzzy c-means clustering models for multivariate time-varying data: Different approaches. Int. J. Uncertain. Fuzz. Knowl. Syst. **12**, 287–326 (2004)
D'Urso, P., Giordani, P.: A least squares approach to principal component analysis for interval valued data. Chemometr. Intell. Lab. Syst. **70**, 179–192 (2004)
D'Urso, P., Giordani, P.: A weighted fuzzy c-means clustering model for symmetric fuzzy data. Comput. Stat. Data Anal. **50**, 1496–1523 (2006)
D'Urso, P., Giordani, P.: A robust fuzzy k-means clustering model for interval valued data. Comput. Stat. **21**, 251–269 (2006)
Esposito Vinzi, V., Lauro, N.C. (eds.): Partial least squares (Special issue). Comput. Stat. Data Anal. **48**(1), 1–220 (2005)
Ferraro, M.B., Coppi, R., González-Rodríguez, G., Colubi, A.: A linear regression model for imprecise response. Int. J. Approx. Reason. **51**, 759–770 (2010)
Galimberti, G., Soffritti, G.: Model-based methods to identify multiple cluster structures in a data set. Comput. Stat. Data Anal. **52**, 520–536 (2007)
Gifi, A.: Nonlinear Multivariate Data Analysis. Wiley, New York (1990)
Gioia, F, Lauro, N.C.: Principal component analysis with interval data. Comput. Stat. **21**, 343–363 (2005)
Giordani, P.: Three-way analysis of imprecise data. J. Multivar. Anal. **101**, 568–582 (2010)
Giordani, P., Kiers, H.A.L.: Principal component analysis of symmetric fuzzy data. Comput. Stat. Data Anal. **45**, 519–548 (2004)
Giordani, P., Kiers, H.A.L.: Three-way component analysis of interval valued data. J. Chemometr. **18**, 253–264 (2004)
Giordano, F., La Rocca, M., Perna C.: Bootstrap variable selection in neural networks regression models. In: Bock, H.-H., Chiodi, M., Mineo A. (eds.) Advances in Multivariate Data Analysis, pp. 109–120. Springer, Heidelberg (2004)
Gordon, A.D., Vichi, M.: Fuzzy partition models for fitting a set of partitions. Psychometrika **66**, 229–247 (2001)
Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2009)
Irpino, A., Tontodonato, V.: Clustering reduced interval data using Hausdorff distance. Comput. Stat. **21**, 271–288 (2006)
Irpino, A., Verde, R.: Dynamic clustering of interval data using a Wasserstein-based distance. Pattern Recognit. Lett. **29**, 1648–1658 (2008)
Kiers, H.A.L., Vicari, D., Vichi, M.: Simultaneous classification and multidimensional scaling with external information. Psychometrika **70**, 433–460 (2005)
Lauro, N.C., D'Ambra, L.: Non-symmetrical exploratory data analysis. Stat. Appl. **4**, 511–529 (1992)
Lauro, N.C., Palumbo, F.: Principal component analysis of interval data: a symbolic data anlysis approach. Comput. Stat. **15**, 73–87 (2000)
Lauro, N.C., Palumbo, F.: Some results and new perspectives in principal component analysis for interval data. In: Book of Short Papers CLADAG, pp. 237–244 (2003)
Lauro, N.C., Giordano, G., Verde, R.: A multidimensional approach to conjoint analysis. Appl. Stoch. Model. Data Anal. **14**, 265–274 (1998)
Laghi, A., Soffritti, G.: A collinearity based hierarchical method to identify clusters of variables. In: Vichi, M., Monari, P., Mignani, S., Montanari, A. (eds.) New Developments in Classification and Data Analysis, pp. 55–62. Springer, Berlin (2005)

Lombardo, R.: Three-way association measure decompositions: the delta index. J. Stat. Plan. Inference **141**, 1789–1799 (2011)
Lombardo, R., Beh, E., D'Ambra, L.: Non-symmetrical correspondence analysis with ordinal variables using orthogonal polymonials. Comput. Stat. Data Anal. **52**, 566–577 (2007)
Lombardo, R., Carlier, A., D'Ambra, L.: Nonsymmetric correspondence analysis for three-way contingency tables. Methodologica **4**, 59–80 (1996)
Miglio, R., Soffritti, G.: Proximity measures between classification trees. In: Bock, H.-H., Chiodi, M., Mineo, A. (eds.) Advances in Multivariate Data Analysis, pp. 27–37. Springer, Heidelberg (2004)
Miglio, R., Soffritti, G.: Simplifying classification trees through consensus methods. In: Vichi, M., Monari, P., Mignani, S., Montanari, A. (eds.) New Developments in Classification and Data Analysis, pp. 31–37. Springer, Berlin (2005)
Perna, C., Giordano, F.: The hidden layer size on feed-forward neural networks: a statistical point of view. Metron Int. J. Stat. **59**, 217–227 (2001)
Piccarreta, R.: A new splitting criterion for classification trees in the ordinal case. In: Bock, H.-H., Chiodi, M., Mineo A. (eds.) Advances in Multivariate Data Analysis, pp. 39–51. Springer, Heidelberg (2004)
Piccarreta, R.: A new measure of nominal-ordinal association. J. Appl. Stat. **28**, 107–120 (2001)
Rizzi, A.: Clustering per le matrici a tre vie. Statistica **49**, 195–208 (1989)
Rizzi, A., Vichi, M.: Representation, synthesis, variability and data preprocessing of a three-way data set. Comput. Stat. Data Anal. **19**, 203–222 (1995)
Rizzi, A., Vichi, M.: Three-way data set analysis. In: Rizzi, A. (ed.) Some Relations Between Matrices and Structures of Multidimensional Data Analysis, pp. 93–166. Giardini Editori e Stampatori, Pisa (1995)
Rocci, R.: Three-mode factor analysis with binary core and orthonormality constraints. J. Ital. Stat. Soc. **1**, 413–422 (1992)
Rocci, R., Bove, G.: Rotation techniques in asymmetric multidimensional scaling. J. Comput. Graph. Stat. **11**, 405–419 (2002)
Rocci, R., Giordani, P.: A weak degeneracy revealing decomposition for the CANDECOMP/PARAFAC model. J. Chemometr. **24**, 57–66 (2010)
Rocci, R., Ten Berge, J.M.F.: Transforming three-way arrays to maximal simplicity. Psychometrika **67**, 351–365 (2002)
Rocci, R., Vichi, M.: Three-mode component analysis with crisp or fuzzy partition of units. Psychometrika **70**, 715–736 (2005)
Rocci, R., Gattone, A., Vichi, M.: A new dimension reduction method: factor discriminant Kmeans. J. Classif. **28**, 210–226 (2011)
Sandri, M. Zuccolotto, P.: Variable selection using random forests. In: Zani, C., Cerioli, A., Riani, M., Vichi, M. (eds.) Data Analysis, Classification and the Forward Search, pp. 263–270. Springer, Heidelberg (2006)
Siciliano, R., Mola, F.: Multivariate data analysis through classification and regression trees. Comput. Stat. Data Anal. **32**, 285–301 (2000)
Siciliano, R., Mooijaart, A.: Three-factor association models for three-way contingency tables. Comput. Stat. Data Anal. **24**, 337–356 (1997)
Simonetti, B., Beh, E.J., D'Ambra, L.: The analysis of dependence for three ways contingency tables with ordinal variables: a case study of patient satisfaction data. J. Appl. Stat. **37**, 91–103 (2010)
Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M., Lauro, N.C.: PLS path modeling. Comput. Stat. Data Anal. **48**, 159–205 (2005)
Timmerman, M., Ceulemans, E., Kiers, H.A.L., Vichi, M.: Factorial and reduced K-means reconsidered. Comput. Stat. Data Anal. **54**, 1858–1871 (2010)
Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1986)
Vicari, D., Vichi, M.: Structural classification analysis of three-way dissimilarity data. J. Classif. **26**, 121–154 (2009)

Vichi, M.: Principal classifications analysis: a method for generating consensus dendograms and its application to three-way data. Comput. Stat. Data Anal. **27**, 311–331 (1998)
Vichi, M., Kiers, H.A.L.: Factorial k-means analysis for two-way data. Comput. Stat. Data Anal. **37**, 49–64 (2001)
Vichi, M., Rocci, R.: Two-mode multipartitioning. Comput. Stat. Data Anal. **52**, 1984–2003 (2008)
Vichi, M., Saporta, G.: Clustering and disjoint principal component analysis. Comput. Stat. Data Anal. **53**, 3194–3208 (2009)
Vichi, M., Rocci, R., Kiers, H.A.L.: Simultaneous component and clustering models for three-way data: Within and between approaches. J. Classif. **24**, 71–98 (2007)