Special Issue on ‘Time series clustering’

Hans-Hermann Bock

Published online: 8 October 2011
© Springer-Verlag 2011

There exists a great variety of clustering methods for multidimensional data points or dissimilarity data, they are standard tools in many statistical packages or data mining systems. More complex data types require special approaches, either by a reduction to classical cases by suitable preprocessing steps, by specifying appropriate dissimilarity measures, or by considering more sophisticated probabilistic clustering models. This Special Issue of ADAC is devoted to the important situation when data have the form of time series or functions where values are observed at different (discrete or continuously varying) time points. Such data are recorded in many domains such as social sciences, meteorology, finance, marketing, medicine, microbiology, etc. The problem is to group together time series with similar shape or structure. In this framework important aspects are, e.g., the interdependence of observation values at different time points, the interest in exploring class-specific periodicity or functional forms, and the large size and dimension of data collections, e.g., in the case of many time points.

This Special ADAC Issue contains five articles on clustering methods for time series which emphasize different aspects. Three of them are concerned with probabilistic approaches, they model the interdependence structure in different ways: class-specific Markov chain modelling, class-specific subspace detection, and class-specific regression over suitably determined time segments. A fourth paper deals with a special dissimilarity approach for financial data, and the last one concentrates on the case of short time series.

- The first article, an invited paper by *Sylvia Frühwirth-Schnatter*, provides a survey on various recent approaches to model-based clustering for panel or longitudinal data, both for continuous as well as categorical observations. The considered time series are Markov chains with class-specific parameter constellations. For

---

H.-H. Bock (✉)
Institute of Statistics, RWTH Aachen University, Aachen, Germany
e-mail: bock@stochastik.rwth-aachen.de

parameter estimation a Bayesian approach is described that uses Markov chain Monte Carlo methods in its computational part. Various criteria are reviewed in order to select the number of clusters. Finally, an application to a panel of marijuana use among teenagers serves as an illustration.

- The next article, written by *Charles Bouveyron* and *Julien Jacques*, develops a general procedure for clustering functional data, as a generalization of a former method HDDC proposed in a multivariate data context. The procedure is based on a functional latent mixture model where data functions are projected into group-specific functional subspaces. By constraining model parameters within and between groups, a family of parsimonious models is exhibited which allows to fit onto various situations. Model parameters and the group-specific functional subspaces are estimated by an EM algorithm. The method and the interpretation of its results is illustrated in the case of meteorological data from Canada, and for four benchmark data sets.
- Apart from clustering time series into groups corresponding to similar shapes or structures, various applications require a segmentation of the time domain into different contiguous time intervals (segments) where the data show a segment-specific, typical behaviour. This problem and corresponding clustering methods are investigated in the paper by *Allou Samé*, *Faïcel Chamroukhi*, *Gérard Govaert*, and *Patrice Aknin*. Within the context of a railway application, they propose a mixture model where clustering and segmentation is combined. Within each segment, the time series are described by a class-specific (e.g., polynomial) regression model, and polynomial coefficients may vary according to a discrete hidden process. The number of clusters and time segments are determined by the classical BIC criterion.
- A broad range of methods for clustering time series proceeds by defining a suitable dissimilarity measure between time series and then applying a classical clustering method to the resulting dissimilarity matrix. The adequate, problem-specific choice of the dissimilarity measure determines largely the success of this approach. In the fourth paper of this issue *Giovanni De Luca* and *Paola Zuccolotto* have applied this two-step strategy to financial time series with a view to a defensive portfolio construction. Their dissimilarity index is defined in terms of the (upper, lower) tail-dependence coefficients of the joint return distribution (of the observed residuals obtained after some preprocessing of the time series), in order to single out pairs of assets which behave similarly in unfavorable market situations. Estimation of these coefficients is performed by using a suitable copula function. Finally it is shown how the clustering results can be used to construct defensive portfolios reducing the effect of an extended financial crisis, for the case of 23 MSCI Developed Market indices.
- The last paper, written by *Tobias Springer*, *Joachim Stöckler*, and *Katja Ickstadt*, starts from a known clustering approach STEM (Short Time Series Expression Miner) for time series that assigns time series to the closest of suitably selected prototypes followed by the selection of significant clusters and eventual grouping. Assuming a correlation-based dissimilarity measure, the paper interprets this procedure in a geometrical way and proposes a modified and (in contrast to STEM) data-dependent criterion for selecting optimum prototypes. This criterion uses

frames (instead of orthogonal bases) in Euclidean space and formulates various connections between frame theory and the considered prototype optimization problem. A practical example illustrates this approach.

The editor gratefully acknowledges the assistance of the following experts and colleagues in the process of reviewing the manuscripts that were submitted for this Special Issue: Wolfgang Aigner (Vienna, Austria), Konstantinos Blekas (Ioannina, Greece), Andrea Cerioli (Parma, Italy), Eva Ceulemans (Leuven, Belgium), Siew Ann Cheong (Singapore), Marco Cuturi (Kyoto, Japan), Ahlame Douzal (Grenoble, France), Martin Ehler (Bethesda, MD, USA), Jason Ernst (Cambridge, MA, USA), Javier Gonzalez (Groningen, The Netherlands), Frank Höppner (Wolfsburg, Germany), Kostas Kalpakis (Baltimore, MD, USA), Peter Kampstra (Amsterdam, The Netherlands), Richard W. Katz (Boulder, CO, USA), T. Warren Liao (Baton Rouge, LA, USA), Fabrizio Lillo (Pisa, Italy), Jennifer Listgarten (Los Angeles, CA, USA), Hermann Locarek-Junge (Dresden, Germany), Francisco Martínez Álvarez (Sevilla, Spain), Carlos Maté (Madrid, Spain), Mauricio Orozco-Alzate (Manizales, Colombia), Sergio Ortobelli (Bergamo, Italy), Edoardo Otranto (Sassari, Italy), Panagiotis Papapetrou (Espoo, Finland), Raffaella Piccarreta (Milano, Italy), Wolfgang Polasek (Vienna, Austria), Julian Antonio Pucheta (Cordoba, Argentina), Carsten Riggelsen (Potsdam, Germany), Gunter Ritter (Passau, Germany), Cristian Rodriguez Rivero (Cordoba, Argentina), Fabrice Rossi (Paris, France), Volker Roth (Basel, Switzerland), Tobias Rydén (Stockholm, Sweden), Theresa Scharl (Vienna, Austria), Stoyan Stoyanov (Singapore), Costanza Torricelli (Modena, Italy), Michele Tumminello (Pittsburgh, PA, USA), Pierpaolo D'Urso (Rome, Italy), Jeroen K. Vermunt (Tilburg, The Netherlands), Juan M. Vilar (Coruña, Spain), Bo Wang (York, UK), Jarke J. van Wijk (Eindhoven, The Netherlands).