

Studies in Classification, Data Analysis, and Knowledge Organization

Managing Editors

H.-H. Bock, Aachen
W. Gaul, Karlsruhe
M. Vichi, Rome
C. Weihs, Dortmund

Editorial Board

D. Baier, Cottbus
F. Critchley, Milton Keynes
R. Decker, Bielefeld
E. Diday, Paris
M. Greenacre, Barcelona
C.N. Lauro, Naples
J. Meulman, Leiden
P. Monari, Bologna
S. Nishisato, Toronto
N. Ohsumi, Tokyo
O. Opitz, Augsburg
G. Ritter, Passau
M. Schader, Mannheim

More information about this series at <http://www.springer.com/series/1564>

Francesco Mola · Claudio Conversano
Maurizio Vichi
Editors

Classification, (Big) Data Analysis and Statistical Learning

 Springer

Editors

Francesco Mola
Department of Business and Economics
University of Cagliari
Cagliari
Italy

Maurizio Vichi
Department of Statistical Sciences
Sapienza University of Rome
Rome
Italy

Claudio Conversano
Department of Business and Economics
University of Cagliari
Cagliari
Italy

ISSN 1431-8814 ISSN 2198-3321 (electronic)
Studies in Classification, Data Analysis, and Knowledge Organization
ISBN 978-3-319-55707-6 ISBN 978-3-319-55708-3 (eBook)
<https://doi.org/10.1007/978-3-319-55708-3>

Library of Congress Control Number: 2017962105

© Springer International Publishing AG, part of Springer Nature 2018, corrected publication 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume comprises the revised versions of the selected papers presented during CLADAG 2015, the 10th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (SIS) that was held in Santa Margherita di Pula, Cagliari, Italy, during October 2015. The meeting took place under the auspices of the International Federation of Classification Societies (IFCS) and of the Italian Statistical Society (SIS). The local organizer was the Department of Business and Economics of the University of Cagliari, Italy.

During the event, advanced methodological researches in multivariate statistics with a special vocation in Data Analysis and Classification were promoted, and the interchange of ideas in these fields of research had strong support, as well as the dissemination of novel concepts, numerical methods, algorithms, computational, and applied results. The scientific program of the conference included three keynote lectures, an invited session, 10 specialized sessions, 15 solicited sessions, and 15 contributed sessions. All the specialized and solicited sessions were promoted by the members of the Scientific Program Committee.

The papers included in this book were submitted after the end of the conference and passed a double-blind revision process. The table of contents is organized basing on specific macro-topics nowadays characterizing different fields of the analysis of complex data structures and Big Data. Accordingly, the volume is organized into the following parts:

- Part I—Big Data,
- Part II—Social Networks,
- Part III—Exploratory Data Analysis,
- Part IV—Statistical Modeling,
- Part V—Clustering and Classification,
- Part VI—Time Series and Spatial Data, and
- Part VII—Finance and Economics.

The editors would like to express their gratitude to all the young statisticians working for Local Organizing Committee (in particular Dr. Massimo Cannas, Dr. Giulia Contu, Dr. Luca Frigau, and Dr. Farideh Tavazoe) for their enthusiasm in supporting the organization of this event from the very beginning. Their hard work contributed relevantly to the success of the event.

Last but not least, we thank all authors and participants, without whom the conference would not have been possible.

Cagliari, Italy
Cagliari, Italy
Rome, Italy
November 2017

Francesco Mola
Claudio Conversano
Maurizio Vichi

Acknowledgements

The organizing committee of the 10th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG) is grateful for the scientific support of members of the International Federation of Classification Societies (IFCS) and of the Italian Statistical Society (SIS). The successful organization of the event benefited from the support of local organizers from the Department of Business and Economics of the University of Cagliari (Italy) and from the financial support of Fondazione Banco di Sardegna. Authors of this book would like to express their very great appreciation to members of both organizations.



SIS -CLADAG
Classification and Data Analysis Group
of the Italian Statistical Society



International Federation
of Classification Societies



Società Italiana di Statistica



Università degli Studi di Cagliari



Fondazione Banco di Sardegna

Contents

Part I Big Data

From Big Data to Information: Statistical Issues Through a Case Study	3
Serena Signorelli and Silvia Biffignandi	
Enhancing Big Data Exploration with Faceted Browsing	13
Sonia Bergamaschi, Giovanni Simonini and Song Zhu	
Big Data Meet Pharmaceutical Industry: An Application on Social Media Data	23
Caterina Liberati and Paolo Mariani	
Electre Tri Machine Learning Approach to the Record Linkage	31
Valentina Minnetti and Renato De Leone	

Part II Social Networks

Finite Sample Behavior of MLE in Network Autocorrelation Models	43
Michele La Rocca, Giovanni C. Porzio, Maria Prosperina Vitale and Patrick Doreian	
Network Analysis Methods for Classification of Roles	51
Simona Gozzo and Venera Tomaselli	
MCA-Based Community Detection	59
Carlo Drago	

Part III Exploratory Data Analysis

Rank Properties for Centred Three-Way Arrays	69
Casper J. Albers, John C. Gower and Henk A. L. Kiers	

Principal Component Analysis of Complex Data and Application to Climatology	77
Sergio Camiz and Silvia Creta	
Motivations and Expectations of Students' Mobility Abroad: A Mapping Technique	87
Valeria Caviezel, Anna Maria Falzoni and Sebastiano Vitali	
Testing Circular Antipodal Symmetry Through Data Depths	97
Giuseppe Pandolfo, Giovanni Casale and Giovanni C. Porzio	
Part IV Statistical Modeling	
Multivariate Stochastic Downscaling for Semicontinuous Data	107
Lucia Paci, Carlo Trivisano and Daniela Cocchi	
Exploring Italian Students' Performances in the SNV Test: A Quantile Regression Perspective	117
Antonella Costanzo and Domenico Vistocco	
Estimating the Effect of Prenatal Care on Birth Outcomes	127
Emiliano Sironi, Massimo Cannas and Francesco Mola	
Part V Clustering and Classification	
Clustering Upper Level Units in Multilevel Models for Ordinal Data	137
Leonardo Grilli, Agnese Panzera and Carla Rampichini	
Clustering Macroseismic Fields by Statistical Data Depth Functions	145
Claudio Agostinelli, Renata Rotondi and Elisa Varini	
Comparison of Cluster Analysis Approaches for Binary Data	155
Giulia Contu and Luca Frigau	
Classification Models as Tools of Bankruptcy Prediction—Polish Experience	163
Józef Pociecha, Barbara Pawelek, Mateusz Baryła and Sabina Augustyn	
Quality of Classification Approaches for the Quantitative Analysis of International Conflict	173
Adalbert F. X. Wilhelm	
Part VI Time Series and Spatial Data	
P-Splines Based Clustering as a General Framework: Some Applications Using Different Clustering Algorithms	183
Carmela Iorio, Gianluca Frasso, Antonio D'Ambrosio and Roberta Siciliano	

Comparing Multistep Ahead Forecasting Functions for Time Series Clustering 191
 Marcella Corduas and Giancarlo Ragozini

Comparing Spatial and Spatio-temporal FPCA to Impute Large Continuous Gaps in Space 201
 Mariantonietta Ruggieri, Antonella Plaia and Francesca Di Salvo

Part VII Finance and Economics

A Graphical Tool for Copula Selection Based on Tail Dependence 211
 Roberta Pappadà, Fabrizio Durante and Nicola Torelli

Bayesian Networks for Financial Market Signals Detection 219
 Alessandro Greppi, Maria E. De Giuli, Claudia Tarantola and Dennis M. Montagna

A Multilevel Heckman Model to Investigate Financial Assets Among Older People in Europe 227
 Omar Paccagnella and Chiara Dal Bianco

Bifurcation and Sunspots in Continuous Time Optimal Model with Externalities 235
 Beatrice Venturi and Alessandro Pirisinu

Erratum to: Big Data Meet Pharmaceutical Industry: An Application on Social Media Data Learning E1
 Caterina Liberati and Paolo Mariani

Contributors

Claudio Agostinelli Dipartimento di Matematica, Università di Trento, Trento, Italy

Casper J. Albers Department of Psychometrics & Statistics, University of Groningen, Groningen, The Netherlands

Sabina Augustyn Department of Statistics, Cracow University of Economics, Cracow, Poland

Mateusz Baryła Department of Statistics, Cracow University of Economics, Cracow, Poland

Sonia Bergamaschi Department of Engineering “Enzo Ferrari”, Università di Modena e Reggio Emilia, Modena, Italy

Silvia Biffignandi University of Bergamo, Bergamo, Italy

Sergio Camiz Dipartimento di Matematica, Sapienza Università di Roma, Rome, Italy

Massimo Cannas Dipartimento di Scienze Economiche ed Aziendali, Università di Cagliari, Cagliari, Italy

Giovanni Casale Department of Economics and Law, University of Cassino and Southern Lazio, Cassino, Italy

Valeria Caviezel Department of Management, Economics and Quantitative Methods, University of Bergamo, Bergamo, Italy

Daniela Cocchi Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

Giulia Contu Department of Economics and Business, University of Cagliari, Cagliari, Italy

Marcella Corduas Department of Political Sciences, University of Naples Federico II, Naples, Italy

Antonella Costanzo National Institute for the Evaluation of Education System – INVALSI, Roma, Italy

Silvia Creta Sapienza Università di Roma, Rome, Italy

Chiara Dal Bianco Department of Economics and Management, University of Padua, Padova, Italy

Antonio D’Ambrosio Department of Economics and Statistics, University of Naples Federico II, Napoli, Italy

Maria E. De Giuli Department of Economics and Management, University of Pavia, Pavia, Italy

Renato De Leone School of Science and Technology, Section of Mathematics, University of Camerino, Camerino, Italy

Francesca Di Salvo Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italy

Patrick Doreian Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia; Department of Sociology, University of Pittsburgh, Pittsburgh, USA

Carlo Drago University of Rome ‘Niccolo Cusano’, Rome, Italy

Fabrizio Durante Dipartimento di Scienze dell’Economia, Università del Salento, Lecce, Italy

Anna Maria Falzoni Department of Management, Economics and Quantitative Methods, University of Bergamo, Bergamo, Italy

Gianluca Frasso Faculté des Sciences Sociales, University of Liège, Liège, Belgium

Luca Frigau Department of Economics and Business, University of Cagliari, Cagliari, Italy

John C. Gower Department of Mathematics & Statistics, The Open University, Milton Keynes, UK

Simona Gozzo Department of Political and Social Sciences, University of Catania, Catania, Italy

Alessandro Greppi Department of Economics and Management, University of Pavia, Pavia, Italy

Leonardo Grilli Department of Statistics, Computer Science, Applications ‘G. Parenti’ University of Florence, Florence, Italy

Carmela Iorio Department of Industrial Engineering, University of Naples Federico II, Napoli, Italy

Henk A. L. Kiers Department of Psychometrics & Statistics, University of Groningen, Groningen, The Netherlands

Michele La Rocca Department of Economics and Statistics, University of Salerno, Fisciano, Italy

Caterina Liberati DEMS, Università degli Studi di Milano-Bicocca, Milan, Italy

Paolo Mariani DEMS, Università degli Studi di Milano-Bicocca, Milan, Italy

Valentina Minnetti Department of Statistic Science, Faculty of Information Engineering, Informatics and Statistics, Sapienza University of Rome, Rome, Italy

Francesco Mola Dipartimento di Scienze Economiche ed Aziendali, Università di Cagliari, Cagliari, Italy

Dennis M. Montagna Department of Economics and Management, University of Pavia, Pavia, Italy

Omar Paccagnella Department of Statistical Sciences, University of Padua, Padova, Italy

Lucia Paci Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

Giuseppe Pandolfo Department of Industrial Engineering, University of Naples Federico II, Naples, Italy

Agnese Panzera Department of Statistics, Computer Science, Applications ‘G. Parenti’ University of Florence, Florence, Italy

Roberta Pappadà Department of Economics, Business, Mathematics and Statistics, University of Trieste, Trieste, Italy

Barbara Pawelek Department of Statistics, Cracow University of Economics, Cracow, Poland

Alessandro Pirisinu Department of Business and Economics, University of Cagliari, Cagliari, Italy

Antonella Plaia Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italy

Józef Pociecha Department of Statistics, Cracow University of Economics, Cracow, Poland

Giovanni C. Porzio Department of Economics and Law, University of Cassino and Southern Lazio, Cassino, Italy

Giancarlo Ragozini Department of Political Sciences, University of Naples Federico II, Naples, Italy

Carla Rampichini Department of Statistics, Computer Science, Applications ‘G. Parenti’ University of Florence, Florence, Italy

Renata Rotondi CNR—Istituto di Matematica Applicata e Tecnologie Informatiche Enrico Magenes, Milano, Italy

Mariantonietta Ruggieri Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italy

Roberta Siciliano Department of Industrial Engineering, University of Naples Federico II, Napoli, Italy

Serena Signorelli University of Bergamo, Bergamo, Italy

Giovanni Simonini Department of Engineering “Enzo Ferrari”, Università di Modena e Reggio Emilia, Modena, Italy

Emiliano Sironi Dipartimento di Scienze Statistiche, Università Cattolica del Sacro Cuore, Milan, Italy

Claudia Tarantola Department of Economics and Management, University of Pavia, Pavia, Italy

Venera Tomaselli Department of Political and Social Sciences, University of Catania, Catania, Italy

Nicola Torelli Department of Economics, Business, Mathematics and Statistics, University of Trieste, Trieste, Italy

Carlo Trivisano Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

Elisa Varini CNR—Istituto di Matematica Applicata e Tecnologie Informatiche Enrico Magenes, Milano, Italy

Beatrice Venturi Department of Business and Economics, University of Cagliari, Cagliari, Italy

Domenico Vistocco Dipartimento di Economia e Giurisprudenza, Università degli Studi di Cassino e del Lazio Meridionale, Cassino, Frosinone, Italy

Maria Prosperina Vitale Department of Economics and Statistics, University of Salerno, Fisciano, Italy

Sebastiano Vitali Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

Adalbert F. X. Wilhelm Jacobs University Bremen, Bremen, Germany

Song Zhu International ICT Doctorate School, Università di Modena e Reggio Emilia, Modena, Italy