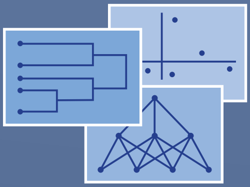
Studies in Classification, Data Analysis, and Knowledge Organization

Theodore Chadjipadelis · Berthold Lausen · Angelos Markos · Tae Rim Lee · Angela Montanari · Rebecca Nugent *Editors*

Data Analysis and Rationality in a Complex World



Springer

Studies in Classification, Data Analysis, and Knowledge Organization

Managing Editors

Wolfgang Gaul, Karlsruhe, Germany

Maurizio Vichi, Rome, Italy

Claus Weihs, Dortmund, Germany

Editorial Board

Daniel Baier, Bayreuth, Germany

Frank Critchley, Milton Keynes, UK

Reinhold Decker, Bielefeld, Germany

Edwin Diday, Paris, France

Michael Greenacre, Barcelona, Spain

Carlo Natale Lauro, Naples, Italy

Jacqueline Meulman, Leiden, The Netherlands

Paola Monari, Bologna, Italy

Shizuhiko Nishisato, Toronto, Canada

Noboru Ohsumi, Tokyo, Japan

Otto Opitz, Augsburg, Germany

Gunter Ritter, Passau, Germany

Martin Schader, Mannheim, Germany

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

Theodore Chadjipadelis · Berthold Lausen ·
Angelos Markos · Tae Rim Lee ·
Angela Montanari · Rebecca Nugent
Editors

Data Analysis and Rationality in a Complex World

Springer

Editors

Theodore Chadjipadelis
Department of Political Sciences
Aristotle University of Thessaloniki
Thessaloniki, Greece

Berthold Lausen
Department of Mathematical Sciences
University of Essex
Colchester, UK

Angelos Markos
School of Education
Democritus University of Thrace
Alexandroupolis, Greece

Tae Rim Lee
Department of Data Science and Statistics
Korea National Open University
Seoul, Korea (Republic of)

Angela Montanari
Department of Statistical Sciences
"Paolo Fortunati"
University of Bologna
Bologna, Italy

Rebecca Nugent
Department of Statistics and Data Science
Carnegie Mellon University
Pittsburgh, PA, USA

ISSN 1431-8814 ISSN 2198-3321 (electronic)
Studies in Classification, Data Analysis, and Knowledge Organization
ISBN 978-3-030-60103-4 ISBN 978-3-030-60104-1 (eBook)
https://doi.org/10.1007/978-3-030-60104-1

Mathematics Subject Classification: 62-XX, 62-06, 62-07, 62Hxx, 62H30, 62Pxx, 62Jxx

© Springer Nature Switzerland AG 2021
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, expressed 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.

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

Preface

This volume contains revised versions of the selected papers presented at the 16th Biennial Conference of the International Federation of Classification Societies (IFCS 2019) organized by the Greek Society of Data Analysis (GSDA), held in Thessaloniki, Greece on 26–29 August 2019. The theme of the conference was "Data Analysis and Rationality in a Complex World". Rationality is a critical issue, as we experience it today. The COVID-19 outbreak revealed also the complexity. Data Analysis is -not the only, but a critical tool for handling information and making decisions under uncertainty on many occasions and for many scientific areas. Rationality is about decision-making based on facts, political and social choice, and the Interest of the People. Authorities, universities, and institutions should take care in order to improve everyday life and solve major political and social problems bringing together Data Science [improve rationality], free and fair Elections [secure free choice and responsibility], and Governance [handling a complex World].

Theodore Chadjipadelis (Aristotle University of Thessaloniki) chaired the Local Organizing Committee and the Scientific Program Committee with Berthold Lausen (IFCS President) and Tae Rim Lee (Korea National Open University) as the vice-chairpersons. The conference encompassed 178 presentations in 56 sessions, including 8 plenary talks and 2 workshops. With 224 attendees from 29 countries, the conference provided a very attractive interdisciplinary international forum for discussion, mutual exchange of knowledge, and cross-disciplinary cooperation.

This volume presents 37 articles dealing with theoretical aspects, methodological advances, and practical applications in domains relating to classification and clustering. The contributions were selected in a second reviewing process after the conference. In addition to the fundamental areas of classification and clustering, the volume contains manuscripts concerning data analysis and statistical modelling in application areas such as economics and finance, computer science, political science, and education. The contributions are listed in alphabetical order with respect to the authors' names.

For the convenience of the reader, the content of this volume is briefly reviewed: *Bellanger et al.* present an agglomerative hierarchical clustering method with temporal ordering constraints. *Chadjipadelis & Teperoglou* employ hierarchical clustering and multiple correspondence analysis to analyze political competition in EU member states at the occasion of the 2019 European Parliament elections. *Champagne Gareau et al.* present a graph clustering technique to improve the efficiency of an electric vehicle planner. *Di Mari et al.* present an approach for computing the coefficient of determination for mixtures of regressions in the Gaussian framework. *Dziechciarz & Dziechciarz-Duda* present a procedure for survey data collection based on fuzzy coding. *Ferreira & Marques* study the relationships between performance measures in discrete supervised classification. *Ganczarek-Gamrot et al.* evaluate value-at-risk measures to assess the risk of price changes in the energy market. *Górecki et al.* define and evaluate measures of mutual dependence for multivariate functional data. *Iodice D'Enza et al.* present a chunk-wise version of iterative principal component analysis for single imputation of "tall" data sets. *Jimeno et al.* run a benchmarking study to evaluate the performance of different clustering methods for mixed-type data. *Kazana et al.* employ a joint dimension reduction and clustering approach to investigate entrepreneurs' attitudes toward a green infrastructure plan. Kitanishi et al. apply a topological data analysis mapper and a spatial perception method to systematically visualize the relationships among pharmaceutical data. *Koutsoupias & Mikelis* combine the use of text mining and multivariate data analysis methods to explore a set of textual documents. *Krężołek & Trzpiot* present an approach to estimate extreme risk using the Hill estimator and its modifications. *Lelu & Cadot* evaluate a series of clustering methods on text data. *Liang & Lee* present experimental results to obtain a rule-of-thumb for choosing the basis spacing for process convolution Gaussian process models. *McLachlan & Ahfock* review and present new results about using the Gaussian mixture model for partially classified data. *Menexes & Koutsos* combine correspondence analysis and ordinary kriging to display values of quantitative variables as supplementary onto factorial maps. *Moschidis & Thanopoulos* apply dimension reduction and clustering techniques to study heterogeneity in e-commerce data from official statistics. *Murugesan et al.* run a benchmarking study to highlight the advantages and drawbacks of spectral clustering, DBSCAN, and k-means on simulated and empirical data. Nakayama employs Bayesian network analysis to model trends in consumer web communication data of new products. *Nicolussi et al.* consider chain graph models for categorical variables to evaluate the level of perceived health in the EU. *Nienkemper-Swanepoel et al.* present a visualization approach to identify the missing data mechanism in incomplete multivariate categorical data. *Okada & Yokoyama* introduce a procedure for assembling one-mode three-way proximities from one-mode two-way proximities, and a method for hierarchical clustering of one-mode three-way proximities. *Panagiotidou & Chadjipadelis* explore the views and attitudes of first-time young voters about Europe and Democracy using multivariate data analysis techniques. *Pratsinakis et al.* compare hierarchical clustering approaches for binary data from molecular markers using external criteria for cluster validation. Smaga introduces permutation and bootstrap tests for the repeated measures analysis of variance for functional data. *Sokołowski & Markowska* present an algorithm for creating a robust distance matrix between observations with outliers. *Srakar & Vecco* present a clustering algorithm for polygonal data. *Stalidis et al.* evaluate the performance of multiple correspondence analysis and hierarchical clustering, as well as a two-layer shallow neural network for personalized supermarket offer recommendations. *Szilágyi & Lengyel* present the results of an empirical study on what motivates the participants of the sharing economy in Hungary using structural equation modeling. *Tai & Frisoli* run a benchmark comparison of minimax linkage to other hierarchical clustering methods using multiple performance metrics on data sets with known clustering structure. *Trejos-Zelaya et al.* implement and evaluate clustering algorithms based on combinatorial optimization metaheuristics. *Tsimperidis et al.* employ keystroke dynamics and machine learning models to classify unknown Internet users according to age, handedness, and educational level. *Varga & Fodor* use hierarchical clustering to derive a typology of critical raw materials with regard to technological innovation. *Vicente-Villardón et al.* extend redundancy analysis to binary data using logistic regression. *Warrens & Ebert* study the predictive power of cluster solutions based on normal mixture models when relevant outcomes are involved in the estimation procedure, using a real-world data set on school motivation.

We would like to express our gratitude to all members of the scientific program committee, for their ability in attracting interesting contributions. A special thanks is due to the local organizing committee for a well-organized conference. We also thank the session organizers for supporting the spread of information about the conference, and for inviting speakers, the reviewers for their timely reports, and Veronika Rosteck and Boopalan Renu of Springer Nature for their support and dedication to the production of this volume. Last but not least, we would like to thank all participants of the conference for their interest and various activities which made the IFCS 2019 conference and this volume an interdisciplinary possibility for scientific discussion.

Colchester, UK	Berthold Lausen
Thessaloniki, Greece	Theodore Chadjipadelis
Alexandroupolis, Greece	Angelos Markos
Seoul, Korea (Republic of)	Tae Rim Lee
Bologna, Italy	Angela Montanari
Pittsburgh, USA	Rebecca Nugent
June 2020	

Contents