Chapman & Hall/CRC

Handbooks of Modern Statistical Methods

Handbook of Cluster Analysis

Edited by

Christian Hennig
Marina Meila
Fionn Murtagh
Roberto Rocci

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Handbook of Cluster Analysis

Chapman & Hall/CRC Handbooks of Modern Statistical Methods

Series Editor

Garrett Fitzmaurice
Department of Biostatistics
Harvard School of Public Health
Boston, MA, U.S.A.

Aims and Scope

The objective of the series is to provide high-quality volumes covering the state-of-the-art in the theory and applications of statistical methodology. The books in the series are thoroughly edited and present comprehensive, coherent, and unified summaries of specific methodological topics from statistics. The chapters are written by the leading researchers in the field, and present a good balance of theory and application through a synthesis of the key methodological developments and examples and case studies using real data.

The scope of the series is wide, covering topics of statistical methodology that are well developed and find application in a range of scientific disciplines. The volumes are primarily of interest to researchers and graduate students from statistics and biostatistics, but also appeal to scientists from fields where the methodology is applied to real problems, including medical research, epidemiology and public health, engineering, biological science, environmental science, and the social sciences.

Published Titles

Handbook of Mixed Membership Models and Their Applications
Edited by Edoardo M. Airoldi, David M. Blei,
Elena A. Erosheva, and Stephen E. Fienberg

Handbook of Markov Chain Monte Carlo
Edited by Steve Brooks, Andrew Gelman,
Galin L. Jones, and Xiao-Li Meng

Handbook of Discrete-Valued Time Series
Edited by Richard A. Davis, Scott H. Holan,
Robert Lund, and Nalini Ravishanker

Handbook of Design and Analysis of Experiments
Edited by Angela Dean, Max Morris,
John Stufken, and Derek Bingham

Longitudinal Data Analysis
Edited by Garrett Fitzmaurice, Marie Davidian,
Geert Verbeke, and Geert Molenberghs

Handbook of Spatial Statistics
Edited by Alan E. Gelfand, Peter J. Diggle,
Montserrat Fuentes, and Peter Guttorp

Handbook of Cluster Analysis
Edited by Christian Hennig, Marina Meila,
Fionn Murtagh, and Roberto Rocci

Handbook of Survival Analysis
Edited by John P. Klein, Hans C. van Houwelingen,
Joseph G. Ibrahim, and Thomas H. Scheike

Handbook of Missing Data Methodology
Edited by Geert Molenberghs, Garrett Fitzmaurice,
Michael G. Kenward, Anastasios Tsiatis, and Geert Verbeke

Chapman & Hall/CRC

Handbooks of Modern Statistical Methods

Handbook of Cluster Analysis

Edited by

Christian Hennig

University College London, UK

Marina Meila

University of Washington, Seattle, USA

Fionn Murtagh

University of Derby, UK

Goldsmiths, University of London, UK

Roberto Rocci

University of Rome Tor Vergata, Italy

CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the

Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20151012

International Standard Book Number-13: 978-1-4665-5189-3 (eBook - PDF)

Contents

Preface xi
Editors xv
Contributors xvii

1. Cluster Analysis: An Overview 1
Christian Hennig and Marina Meila

2. A Brief History of Cluster Analysis 21
Fionn Murtagh

Section I Optimization Methods

3. Quadratic Error and k-Means 33
Boris Mirkin

4. *K*-Medoids and Other Criteria for Crisp Clustering 55
Douglas Steinley

5. Foundations for Center-Based Clustering: Worst-Case Approximations and Modern Developments 67
Pranjal Awasthi and Maria Florina Balcan

Section II Dissimilarity-Based Methods

6. Hierarchical Clustering 103
Pedro Contreras and Fionn Murtagh

7. Spectral Clustering 125
Marina Meila

Section III Methods Based on Probability Models

8. Mixture Models for Standard *p*-Dimensional Euclidean Data 145
Geoffrey J. McLachlan and Suren I. Rathnayake

9. Latent Class Models for Categorical Data 173
G. Celeux and Gérard Govaert

10. Dirichlet Process Mixtures and Nonparametric Bayesian Approaches to Clustering 195
Vinayak Rao

11. Finite Mixtures of Structured Models 217
Marco Alfó and Sara Viviani

12. Time-Series Clustering 241
Jorge Caiado, Elizabeth Ann Maharaj, and Pierpaolo D'Urso

13. Clustering Functional Data 265
David B. Hitchcock and Mark C. Greenwood

14. Methods Based on Spatial Processes 289
Lisa Handl, Christian Hirsch, and Volker Schmidt

15. Significance Testing in Clustering 315
Hanwen Huang, Yufeng Liu, David Neil Hayes, Andrew Nobel, J.S. Marron, and Christian Hennig

16. Model-Based Clustering for Network Data 337
Thomas Brendan Murphy

Section IV Methods Based on Density Modes and Level Sets

17. A Formulation in Modal Clustering Based on Upper Level Sets 361
Adelchi Azzalini

18. Clustering Methods Based on Kernel Density Estimators: Mean-Shift Algorithms 383
Miguel Á. Carreira-Perpiñán

19. Nature-Inspired Clustering 419
Julia Handl and Joshua Knowles

Section V Specific Cluster and Data Formats

20. Semi-Supervised Clustering 443
Anil Jain, Rong Jin, and Radha Chitta

21. Clustering of Symbolic Data 469
Paula Brito

22. A Survey of Consensus Clustering 497
Joydeep Ghosh and Ayan Acharya

23. Two-Mode Partitioning and Multipartitioning 519
Maurizio Vichi

24. Fuzzy Clustering 545
Pierpaolo D'Urso

25. Rough Set Clustering 575
Ivo Düntsch and Günther Gediga

Section VI Cluster Validation and Further General Issues

26. Method-Independent Indices for Cluster Validation and Estimating the Number of Clusters 595
Maria Halkidi, Michalis Vazirgiannis, and Christian Hennig

27. Criteria for Comparing Clusterings 619
Marina Meila

28. Resampling Methods for Exploring Cluster Stability 637
Friedrich Leisch

29. Robustness and Outliers 653
L.A. García-Escudero, A. Gordaliza, C. Matrán, A. Mayo-Iscar, and Christian Hennig

30. Visual Clustering for Data Analysis and Graphical User Interfaces 679
Sébastien Déjean and Josiane Mothe

31. Clustering Strategy and Method Selection 703
Christian Hennig