

---

# Contents

---

## Part I New Methods in Cluster Analysis

---

### Thinking Ultrametrically

*Fionn Murtagh* ..... 3

### Clustering by Vertex Density in a Graph

*Alain Guénoche* ..... 15

### Clustering by Ant Colony Optimization

*Javier Trejos, Alex Murillo, Eduardo Piza* ..... 25

### A Dynamic Cluster Algorithm Based on $L_r$ Distances for Quantitative Data

*Francisco de A. T. de Carvalho, Yves Lechevallier, Renata M.C.R. de Souza* ..... 33

### The Last Step of a New Divisive Monothetic Clustering Method: the Gluing-Back Criterion

*Jean-Yves Pirçon, Jean-Paul Rasson* ..... 43

### Standardizing Variables in $K$ -means Clustering

*Douglas Steinley* ..... 53

### A Self-Organizing Map for Dissimilarity Data

*Aïcha El Golli, Brieuc Conan-Guez, Fabrice Rossi* ..... 61

### Another Version of the Block EM Algorithm

*Mohamed Nadif, Gérard Govaert* ..... 69

### Controlling the Level of Separation of Components in Monte Carlo Studies of Latent Class Models

*José G. Dias* ..... 77

**Fixing Parameters in the Constrained Hierarchical Classification Method: Application to Digital Image Segmentation**

*Kaddour Bachar, Israël-César Lerman* ..... 85

**New Approaches for Sum-of-Diameters Clustering**

*Sarnath Ramnath, Mynul H. Khan,, Zubair Shams* ..... 95

**Spatial Pyramidal Clustering Based on a Tessellation**

*Edwin Diday* ..... 105

---

**Part II Modern Nonparametrics**

---

**Relative Projection Pursuit and its Application**

*Masahiro Mizuta and Shintaro Hiro* ..... 123

**Priors for Neural Networks**

*Herbert K. H. Lee* ..... 141

**Combining Models in Discrete Discriminant Analysis Through a Committee of Methods**

*Ana Sousa Ferreira* ..... 151

**Phoneme Discrimination with Functional Multi-Layer Perceptrons**

*Brieuc Conan-Guez, Fabrice Rossi* ..... 157

**PLS Approach for Clusterwise Linear Regression on Functional Data**

*Cristian Preda, Gilbert Saporta* ..... 167

**On Classification and Regression Trees for Multiple Responses**

*Seong Keon Lee* ..... 177

**Subsetting Kernel Regression Models Using Genetic Algorithm and the Information Measure of Complexity**

*Xinli Bao, Hamparsum Bozdogan* ..... 185

**Cherry-Picking as a Robustness Tool**

*Leanna L. House, David Banks* ..... 197

---

**Part III Classification and Dimension Reduction**

---

**Academic Obsessions and Classification Realities: Ignoring Practicalities in Supervised Classification**

*David J. Hand* ..... 209

<b>Modified Biplots for Enhancing Two-Class Discriminant Analysis</b>	
<i>Sugnet Gardner, Niel le Roux</i> .....	233
<b>Weighted Likelihood Estimation of Person Locations in an Unfolding Model for Polytomous Responses</b>	
<i>Guanzhong Luo, David Andrich</i> .....	241
<b>Classification of Geospatial Lattice Data and their Graphical Representation</b>	
<i>Koji Kurihara</i> .....	251
<b>Degenerate Expectation-Maximization Algorithm for Local Dimension Reduction</b>	
<i>Xiaodong Lin, Yu Zhu</i> .....	259
<b>A Dimension Reduction Technique for Local Linear Regression</b>	
<i>Joseph Lucas</i> .....	269
<b>Reducing the Number of Variables Using Implicative Analysis</b>	
<i>Raphaël Couturier, Régis Gras, Fabrice Guillet</i> .....	277
<b>Optimal Discretization of Quantitative Attributes for Association Rules</b>	
<i>Stefan Born, Lars Schmidt-Thieme</i> .....	287
<hr/> <b>Part IV Symbolic Data Analysis</b> <hr/>	
<b>Clustering Methods in Symbolic Data Analysis</b>	
<i>Rosanna Verde</i> .....	299
<b>Dependencies in Bivariate Interval-Valued Symbolic Data</b>	
<i>L. Billard</i> .....	319
<b>Clustering of Symbolic Objects Described by Multi-Valued and Modal Variables</b>	
<i>André Hardy, Pascale Lallemand</i> .....	325
<b>A Hausdorff Distance Between Hyper-Rectangles for Clustering Interval Data</b>	
<i>Marie Chavent</i> .....	333
<b>Kolmogorov-Smirnov for Decision Trees on Interval and Histogram Variables</b>	
<i>Chérif Mballo and Edwin Diday</i> .....	341

**Dynamic Cluster Methods for Interval Data Based on Mahalanobis Distances**

*Renata M.C.R. de Souza, Francisco de A.T. de Carvalho, Camilo P. Tenório, Yves Lechevallier* ..... 351

**A Symbolic Model-Based Approach for Making Collaborative Group Recommendations**

*Sérgio R. de M. Queiroz, Francisco de A. T. de Carvalho* ..... 361

**Probabilistic Allocation of Aggregated Statistical Units in Classification Trees for Symbolic Class Description**

*Mohamed Mehdi Limam, Edwin Diday, Suzanne Winsberg* ..... 371

**Building Small Scale Models of Multi-Entity Databases by Clustering**

*Georges Hébrail, Yves Lechevallier* ..... 381

---

**Part V Taxonomy and Medicine**

---

**Phylogenetic Closure Operations and Homoplasy-Free Evolution**

*Tobias Dezulian, Mike Steel* ..... 395

**Consensus of Classification Systems, with Adams' Results Revisited**

*Florent Domenach, Bruno Leclerc* ..... 417

**Symbolic Linear Regression with Taxonomies**

*F. Afonso, L. Billard, and E. Diday* ..... 429

**Determining Horizontal Gene Transfers in Species Classification: Unique Scenario**

*Vladimir Makarenkov, Alix Boc, Abdoulaye Baniré Diallo* ..... 439

**Active and Passive Learning to Explore a Complex Metabolism Data Set**

*Susan J. Simmons, Xiaodong Lin, Chris Beecher, Young Truong and S. Stanley Young*..... 447

**Mathematical and Statistical Modeling of Acute Inflammation**

*Gilles Clermont, Carson C. Chow, Gregory M. Constantine, Yoram Vodovotz, John Bartels* ..... 457

**Combining Functional MRI Data on Multiple Subjects**

*Mark G. Vangel*..... 469

**Classifying the State of Parkinsonism by Using Electronic Force Platform Measures of Balance**

*Nicholas I. Bohnen, Marius G. Buliga, Gregory M. Constantine* ..... 477

**Subject Filtering for Passive Biometric Monitoring**

*Vahan Grigoryan, Donald Chiarulli, Milos Hauskrecht* ..... 485

---

**Part VI Text Mining**

---

**Mining Massive Text Data and Developing Tracking Statistics**

*Daniel R. Jeske, Regina Y. Liu* ..... 495

**Contributions of Textual Data Analysis to Text Retrieval**

*Simona Balbi, Emilio Di Meglio* ..... 511

**Automated Resolution of Noisy Bibliographic References**

*Markus Demleitner, Michael Kurtz, Alberto Accomazzi, Günther Eichhorn, Carolyn S. Grant, Steven S. Murray* ..... 521

**Choosing the Right Bigrams for Information Retrieval**

*Maojin Jiang, Eric Jensen, Steve Beitzel, Shlomo Argamon* ..... 531

**A Mixture Clustering Model for Pseudo Feedback in Information Retrieval**

*Tao Tao, ChengXiang Zhai* ..... 541

**Analysis of Cross-Language Open-Ended Questions Through MFACT**

*Mónica Bécue, Jérôme Pagès, Campo-Elías Pardo* ..... 553

**Inferring User's Information Context from User Profiles and Concept Hierarchies**

*Ahu Sieg, Bamshad Mobasher, Robin Burke* ..... 563

**Database Selection for Longer Queries**

*Wensheng Wu, Clement Yu, Weiyi Meng* ..... 575

---

**Part VII Contingency Tables and Missing Data**

---

**An Overview of Collapsibility**

*Stefano De Cantis and Antonino M. Oliveri* ..... 587

**Generalized Factor Analyses for Contingency Tables**

*François Bavaud* ..... 597

<b>A PLS Approach to Multiple Table Analysis</b> <i>Michel Tenenhaus</i> .....	607
<b>Simultaneous Row and Column Partitioning in Several Contingency Tables</b> <i>Vincent Loonis</i> .....	621
<b>Missing Data and Imputation Methods in Partition of Variables</b> <i>Ana Lorga da Silva, Gilbert Saporta, Helena Bacelar-Nicolau</i> .....	631
<b>The Treatment of Missing Values and its Effect on Classifier Accuracy</b> <i>Edgar Acuña, Caroline Rodriguez</i> .....	639
<b>Clustering with Missing Values: No Imputation Required</b> <i>Kiri Wagstaff</i> .....	649

Classification, Clustering, and Data Mining Applications  
Proceedings of the Meeting of the International  
Federation of Classification Societies (IFCS), Illinois  
Institute of Technology, Chicago, 15–18 July 2004  
Banks, D.; House, L.; McMorris, F.R.; Arabie, P.; Gaul, W.  
(Eds.)  
2004, XIV, 658 p., Softcover  
ISBN: 978-3-540-22014-5