

# Contents

---

## Part I. Analysis of Symbolic Data

---

<b>Dependencies and Variation Components of Symbolic Interval-Valued Data</b> .....	3
<i>Lynne Billard</i>	
<b>On the Analysis of Symbolic Data</b> .....	13
<i>Paula Brito</i>	
<b>Symbolic Analysis to Learn Evolving CyberTraffic</b> .....	23
<i>Costantina Caruso, Donato Malerba</i>	
<b>A Clustering Algorithm for Symbolic Interval Data Based on a Single Adaptive Hausdorff Distance</b> .....	35
<i>Francisco de A.T. de Carvalho</i>	
<b>An Agglomerative Hierarchical Clustering Algorithm for Improving Symbolic Object Retrieval</b> .....	45
<i>Floriana Esposito, Claudia d'Amato</i>	
<b>3WAYSYM-SCAL: Three-Way Symbolic Multidimensional Scaling</b>	55
<i>Patrick J.F. Groenen, Suzanne Winsberg</i>	
<b>Clustering and Validation of Interval Data</b> .....	69
<i>André Hardy, Joffray Baune</i>	
<b>Building Symbolic Objects from Data Streams</b> .....	83
<i>Georges Hébrail, Yves Lechevallier</i>	
<b>Feature Clustering Method to Detect Monotonic Chain Structures in Symbolic Data</b> .....	95
<i>Manabu Ichino</i>	
<b>Symbolic Markov Chains</b> .....	103
<i>Monique Noirhomme-Fraiture, Etienne Cuvelier</i>	
<b>Quality Issues in Symbolic Data Analysis</b> .....	113
<i>Haralambos Papageorgiou, Maria Vardaki</i>	
<b>Dynamic Clustering of Histogram Data: Using the Right Metric</b> .....	123
<i>Rosanna Verde, Antonio Irpino</i>	

---

**Part II. Clustering Methods**

---

**Beyond the Pyramids: Rigid Clustering Systems** . . . . . 137  
*Jean-Pierre Barthélemy, Gentian Gusho, Christophe Osswald*

**Indirect Blockmodeling of 3-Way Networks** . . . . . 151  
*Vladimir Batagelj, Anuška Ferligoj, Patrick Doreian*

**Clustering Methods: A History of  $k$ -Means Algorithms** . . . . . 161  
*Hans-Hermann Bock*

**Overlapping Clustering in a Graph Using  $k$ -Means and Application to Protein Interactions Networks** . . . . . 173  
*Irène Charon, Lucile Denœud, Olivier Hudry*

**Species Clustering via Classical and Interval Data Representation** . . . . . 183  
*Marie Chavent*

**Looking for High Density Zones in a Graph** . . . . . 193  
*Tristan Colombo, Alain Guénoche*

**Block Bernoulli Parsimonious Clustering Models** . . . . . 203  
*Gérard Govaert, Mohamed Nadif*

**Cluster Analysis Based on Posets** . . . . . 213  
*Melvin F. Janowitz*

**Hybrid  $k$ -Means: Combining Regression-Wise and Centroid-Based Criteria for QSAR** . . . . . 225  
*Robert Stanforth, Evgueni Kolossov, Boris Mirkin*

**Partitioning by Particle Swarm Optimization** . . . . . 235  
*Javier Trejos-Zelaya, Mario Villalobos-Arias*

---

**Part III. Conceptual Analysis of Data**

---

**Concepts of a Discrete Random Variable** . . . . . 247  
*Richard Emilion*

**Mining Description Logics Concepts with Relational Concept Analysis** . . . . . 259  
*Marianne Huchard, Amedeo Napoli, Mohamed Rouane Hacene, Petko Valtchev*

**Representation of Concept Description by Multivalued Taxonomic Preordnance Variables** ..... 271  
*Israël-César Lerman, Philippe Peter*

**Recent Advances in Conceptual Clustering: CLUSTER3** ..... 285  
*Ryszard S. Michalski, William D. Seeman*

**Symbolic Dynamics in Text: Application to Automated Construction of Concept Hierarchies** ..... 299  
*Fionn Murtagh*

**Part IV. Consensus Methods**

**Average Consensus and Infinite Norm Consensus: Two Methods for Ultrametric Trees** ..... 309  
*Guy Cucumel*

**Consensus from Frequent Groupings** ..... 317  
*Bruno Leclerc*

**Consensus of Star Tree Hypergraphs** ..... 325  
*Fred R. McMorris, Robert C. Powers*

**Part V. Data Analysis, Data Mining, and KDD**

**Knowledge Management in Environmental Sciences with *IKBS*: Application to Systematics of Corals of the Mascarene Archipelago** ..... 333  
*Noel Conruyt, David Grosser*

**Unsupervised Learning Informational Limit in Case of Sparsely Described Examples** ..... 345  
*Jean-Gabriel Ganascia, Julien Velcin*

**Data Analysis and Operations Research** ..... 357  
*Wolfgang Gaul*

**Reduction of Redundant Rules in Statistical Implicative Analysis** ..... 367  
*Régis Gras, Pascale Kuntz*

**Mining Personal Banking Data to Detect Fraud** ..... 377  
*David J. Hand*

**Finding Rules in Data** ..... 387  
*Tu-Bao Ho*

**Mining Biological Data Using Pyramids** ..... 397  
*Géraldine Polaillon, Laure Vescovo, Magali Michaut,  
 Jean-Christophe Aude*

**Association Rules for Categorical and Tree Data** ..... 409  
*Henri Ralambondrainy, Jean Diatta*

**Induction Graphs for Data Mining** ..... 419  
*Djamel Abdelkader Zighed*

---

**Part VI. Dissimilarities: Structures and Indices**

---

**Clustering of Molecules: Influence of the Similarity Measures** . 433  
*Samia Aci, Gilles Bisson, Sylvaine Roy, Samuel Wieczorek*

**Group Average Representations in Euclidean Distance Cones** . 445  
*Casper J. Albers, Frank Critchley, John C. Gower*

**On Lower-Maximal Paired-Ultrametrics** ..... 455  
*Patrice Bertrand, François Brucker*

**A Note on Three-Way Dissimilarities and Their Relationship  
 with Two-Way Dissimilarities** ..... 465  
*Victor Chepoi, Bernard Fichet*

**One-to-One Correspondence Between Indexed Cluster  
 Structures and Weakly Indexed Closed Cluster Structures** .... 477  
*Jean Diatta*

**Adaptive Dissimilarity Index for Gene Expression Profiles  
 Classification** ..... 483  
*Ahlame Douzal Chouakria, Alpha Diallo, Françoise Giroud*

**Lower (Anti-) Robinson Rank Representations for Symmetric  
 Proximity Matrices** ..... 495  
*Lawrence J. Hubert, Hans-Friedrich Köhn*

**Density-Based Distances: a New Approach for Evaluating  
 Proximities Between Objects. Applications in Clustering and  
 Discriminant Analysis** ..... 505  
*Jean-Paul Rasson, François Roland*

**Robinson Cubes** ..... 515  
*Matthijs J. Warrens, Willem J. Heiser*

---

**Part VII. Multivariate Statistics**

---

**Relative and Absolute Contributions to Aid Strata Interpretation** ..... 527  
*M. Carmen Bravo, José M. García-Santesmases*

**Classification and Generalized Principal Component Analysis** . 539  
*Henri Caussinus, Anne Ruiz-Gazen*

**Locally Linear Regression and the Calibration Problem for Micro-Array Analysis** ..... 549  
*Antonio Ciampi, Benjamin Rich, Alina Dyachenko, Isadora Antoniano Villalobos, Carl Murie, Robert Nadon*

**Sanskrit Manuscript Comparison for Critical Edition and Classification** ..... 557  
*Marc Csernel, Patrice Bertrand*

**Divided Switzerland** ..... 567  
*Yadolah Dodge, Gérard Geiser, Valentin Rousson*

**Prediction with Confidence** ..... 577  
*Alexander Gammerman*

**Which Bootstrap for Principal Axes Methods?** ..... 581  
*Ludovic Lebart*

**PCR and PLS for Clusterwise Regression on Functional Data** 589  
*Cristian Preda, Gilbert Saporta*

**A New Method for Ranking  $n$  Statistical Units** ..... 599  
*Alfredo Rizzi*

**About Relational Correlations** ..... 609  
*Yves Schektman*

**Dynamic Features Extraction in Soybean Futures Market of China**..... 619  
*Huiwen Wang, Jie Meng*

**Index** ..... 629