Contents

Part I (Semi-) Plenary Presentations

Hierarchical Clustering with Performance Guarantees 3
Sanjoy Dasgupta

Alignment Free String Distances for Phylogeny 15
Frédéric Guyon and Alain Guénoche

Data Quality Dependent Decision Making in Pattern Classification 25
Josef Kittler and Norman Poh

Clustering Proteins and Reconstructing Evolutionary Events 37
Boris Mirkin

Microarray Dimension Reduction Based on Maximizing Mantel Correlation Coefficients Using a Genetic Algorithm Search Strategy 49
Elena Deych, Robert Culverhouse, and William D. Shannon

Part II Classification and Data Analysis

Classification

Multiparameter Hierarchical Clustering Methods 63
Gunnar Carlsson and Facundo Mémoli

Unsupervised Sparsification of Similarity Graphs 71
Tim Gollub and Benno Stein

Simultaneous Clustering and Dimensionality Reduction Using Variational Bayesian Mixture Model 81
Kazuho Watanabe, Shotaro Akaho, Shinichiro Omachi, and Masato Okada

Bibliografische Informationen
http://d-nb.info/999263552

digitalisiert durch DEUTSCHE NATIONAL BIBLIOTHEK

A Partitioning Method for the Clustering of Categorical Variables 91
Marie Chavent, Vanessa Kuentz, and Jérôme Saracco

Treed Gaussian Process Models for Classification 101
Tamara Broderick and Robert B. Gramacy

Ridgeline Plot and Clusterwise Stability as Tools for Merging Gaussian Mixture Components 109
Christian Hennig

Clustering with Confidence: A Low-Dimensional Binning Approach 117
Rebecca Nugent and Werner Stuetzle

Local Classification of Discrete Variables by Latent Class Models 127
Michael Bücker, Gero Szepannek, and Claus Weihs

A Comparative Study on Discrete Discriminant Analysis through a Hierarchical Coupling Approach 137
Ana Sousa Ferreira

A Comparative Study of Several Parametric and Semiparametric Approaches for Time Series Classification 147
Sonia Pértega Díaz and José A. Vilar

Finite Dimensional Representation of Functional Data with Applications 157
Alberto Muñoz and Javier González

Clustering Spatio-Functional Data: A Model Based Approach 167
Elvira Romano, Antonio Balzanella, and Rosanna Verde

Use of Mixture Models in Multiple Hypothesis Testing with Applications in Bioinformatics 177
Geoffrey J. McLachlan and Leesa Wockner

Finding Groups in Ordinal Data: An Examination of Some Clustering Procedures 185
Marek Walesiak and Andrzej Dudek

An Application of One-mode Three-way Overlapping Cluster Analysis 193
Satoru Yokoyama, Atsuho Nakayama, and Akinori Okada

Evaluation of Clustering Results: The Trade-off Bias-Variability 201
Margarida G.M.S. Cardoso, Katti Faceli, and André C.P.L.F. de Carvalho

Cluster Structured Multivariate Probability Distribution with Uniform Marginals 209
Andrzej Sokolowski and Sabina Denkowska

Analysis of Diversity-Accuracy Relations in Cluster Ensemble 217
Dorota Rozmus

Linear Discriminant Analysis with more Variables than Observations: A not so Naive Approach 227
A. Pedro Duarte Silva

Fast Hierarchical Clustering from the Baire Distance 235
Pedro Contreras and Fionn Murtagh

Data Analysis

The Trend Vector Model: Identification and Estimation in SAS 245
Mark de Rooij and Hsiu-Ting Yu

Discrete Beta-Type Models 253
Antonio Punzo

The R Package DAKS: Basic Functions and Complex Algorithms in Knowledge Space Theory 263
Anatol Sargin and Ali Ünlü

Methods for the Analysis of Skew-Symmetry in Asymmetric Multidimensional Scaling 271
Giuseppe Bove

Canonical Correspondence Analysis in Social Science Research 279
Michael Greenacre

Exploring Data Through Archetypes 287
Maria Rosaria D'Esposito, Giancarlo Ragozini, and Domenico Vistocco

Exploring Sensitive Topics: Sensitivity, Jeopardy, and Cheating 299
Claudia Becker

Sampling the Join of Streams 307
Raphaël Féraud, Fabrice Clérot, and Pascal Gouzien

The R Package fechner for Fechnerian Scaling 315
Thomas Kiefer, Ali Ünlü, and Ehtibar N. Dzhafarov

Asymptotic Behaviour in Symbolic Markov Chains 323
Monique Noirhomme-Fraiture

An Interactive Graphical System for Visualizing Data Quality–Tableplot Graphics 331
Waqas Ahmed Malik, Antony Unwin, and Alexander Gribov

Symbolic Multidimensional Scaling Versus Noisy Variables and Outliers 341
Marcin Pełka

Principal Components Analysis for Trapezoidal Fuzzy Numbers 351
Alexia Pacheco and Oldemar Rodríguez

Factor Selection in Observational Studies – An Application of Nonlinear Factor Selection to Propensity Scores 361
Stephan Dlugosz

Nonlinear Mapping Using a Hybrid of PARAMAP and Isomap Approaches 371
Ulas Akkucuk and J. Douglas Carroll

Dimensionality Reduction Techniques for Streaming Time Series: A New Symbolic Approach 381
Antonio Balzanella, Antonio Irpino, and Rosanna Verde

A Batesian Semiparametric Generalized Linear Model with Random Effects Using Dirichlet Process Priors 391
Kei Miyazaki and Kazuo Shigemasu

Exact Confidence Intervals for Odds Ratios with Algebraic Statistics 399
Anne Krampe and Sonja Kuhnt

The CHIC Analysis Software v1.0 409
Angelos Markos, George Menexes, and Iannis Papadimitriou

Part III Applications

Archaeology and Spatial Planning

Clustering the Roman Heaven: Uncovering the Religious Structures in the Roman Province Germania Superior 419
Tudor Ionescu and Leif Scheuermann

Geochemical and Statistical Investigation of Roman Stamped Tiles of the *Legio XXI Rapax* 427
Hans-Georg Bartel, Hans-Joachim Mucha, and Jens Dolata

Land Cover Classification by Multisource Remote Sensing: Comparing Classifiers for Spatial Data 435
Alexander Brenning

Are there Cluster of Communities with the Same Dynamic Behaviour? 445
Martin Behnisch and Alfred Ultsch

Land Cover Detection with Unsupervised Clustering and Hierarchical Partitioning 455
Laura Poggio and Pierre Soille

Using Advanced Regression Models for Determining Optimal Soil Heterogeneity Indicators 463
Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner

Bio-Sciences

Local Analysis of SNP Data 473
Tina Müller, Julia Schiffner, Holger Schwender, Gero Szepannek, Claus Weihs, and Katja Ickstadt

Airborne Particulate Matter and Adverse Health Events: Robust Estimation of Timescale Effects 481
Massimo Bilancia and Francesco Campobasso

Identification of Specific Genomic Regions Responsible for the Invasivity of *Neisseria Meningitidis* 491
Dunarel Badescu, Abdoulaye Baniré Diallo, and Vladimir Makarenkov

Classification of ABC Transporters Using Community Detection 501
Claire Gaugain, Roland Barriot, Gwennaele Fichant, and Yves Quentin

Estimation of the Number of Sustained Viral Responders by Interferon Therapy Using Random Numbers with a Logistic Model 509
Shinobu Tatsunami, Takahiko Ueno, Rie Kuwabara, Junichi Mimaya, Akira Shirahata, and Masashi Taki

Virtual High Throughput Screening Using Machine Learning Methods 517
Cherif Mballo and Vladimir Makarenkov

Electronic Data and Web

Network Analysis of Works on Clustering and Classification from Web of Science 525
Nataša Kejžar, Simona Korenjak Černe, and Vladimir Batagelj

Recommending in Social Tagging Systems Based on Kernelized Multiway Analysis 537
Alexandros Nanopoulos and Artus Krohn-Grimberghe

Dynamic Population Segmentation in Online Market Monitoring 545
Norbert Walchhofer, Karl A. Froeschl, Milan Hronsky, and Kurt Hornik

Gaining 'Consumer Insights' from Influential Actors in Weblog Networks 553
Martin Klaus and Ralf Wagner

Visualising a Text with a Tree Cloud 561
Philippe Gambette and Jean Véronis

A Tree Kernel Based on Classification and Citation Data to Analyse Patent Documents 571
Markus Arndt and Ulrich Arndt

A New SNA Centrality Measure Quantifying the Distance to the Nearest Center 579
Angela Bohn, Stefan Theußl, Ingo Feinerer, Kurt Hornik, Patrick Mair, and Norbert Walchhofer

Mining Innovative Ideas to Support New Product Research and Development 587
Dirk Thorleuchter, Dirk Van den Poel, and Anita Prinzie

Finance and Banking

The Basis of Credit Scoring: On the Definition of Credit Default Events 595
Alexandra Schwarz and Gerhard Arminger

Forecasting Candlesticks Time Series with Locally Weighted Learning Methods 603
Javier Arroyo

An Analysis of Alternative Methods for Measuring Long-Run Performance: An Application to Share Repurchase Announcements 613
Wolfgang Bessler, Julian Holler, and Martin Seim

Knowledge Discovery in Stock Market Data 621
Alfred Ultsch and Hermann Locarek-Junge

The Asia Financial Crises and Exchange Rates: Had there been Volatility Shifts for Asian Currencies? 629
Takashi Oga and Wolfgang Polasek

The Pricing of Risky Securities in a Fuzzy Least Square Regression Model 639
Francesco Campobasso, Annarita Fanizzi, and Massimo Bilancia

Linguistics

Classification of the Indo-European Languages Using a Phylogenetic Network Approach 647
Alix Boc, Anna Maria Di Sciullo, and Vladimir Makarenkov

Parsing as Classification 657
Lidia Khmylko and Wolfgang Menzel

Comparing the Stability of Clustering Results of Dialect Data Based on Several Distance Matrices 665
Edgar Haimerl and Hans-Joachim Mucha

Marketing

Marketing and Regional Sales: Evaluation of Expenditure Strategies by Spatial Sales Response Functions 673
Daniel Baier and Wolfgang Polasek

A Demand Learning Data Based Approach to Optimize Revenues of a Retail Chain 683
Wolfgang Gaul and Abdolhadi Darzian Azizi

Missing Values and the Consistency Problem Concerning AHP Data 693
Wolfgang Gaul and Dominic Gastes

Monte Carlo Methods in the Assessment of New Products: A Comparison of Different Approaches 701
Said Esber and Daniel Baier

Preference Analysis and Product Design in Markets for Elderly People: A Comparison of Methods and Approaches 709
Samah Abu-Assab, Daniel Baier, and Mirko Kühne

Usefulness of A Priori Information about Customers for Market Research: An Analysis for Personalisation Aspects in Retailing 717
Michael Brusch and Eva Stüber

Importance of Consumer Preferences on the Diffusion of Complex Products and Systems 725
Sabine Schmidt and Magdalena Missler-Behr

Household Possession of Consumer Durables on Background of some Poverty Lines 735
Józef Dziechciarz, Marta Dziechciarz, and Klaudia Przybysz

Effect of Consumer Perceptions of Web Site Brand Personality and Web Site Brand Association on Web Site Brand Image 743
Sandra Loureiro and Silvina Santana

Music Science

Perceptually Based Phoneme Recognition in Popular Music 751
Gero Szepannek, Matthias Gruhne, Bernd Bischl, Sebastian Krey, Tamas Harczos, Frank Klefenz, Christian Dittmar, and Claus Weihs

SVM Based Instrument and Timbre Classification 759
Sebastian Krey and Uwe Ligges

Three-way Scaling and Clustering Approach to Musical Structural Analysis 767
Mitsuhiro Tsuji, Toshio Shimokawa, and Akinori Okada

Improving GMM Classifiers by Preliminary One-class SVM Outlier Detection: Application to Automatic Music Mood Estimation 775
Hanna Lukashevich and Christian Dittmar

Quality Assurance and Engineering

Multiobjective Optimization for Decision Support in Automated 2.5D System-in-Package Electronics Design 783
Martin Berger, Michael Schröder, and Karl-Heinz Küfer

Multi-Objective Quality Assessment for EA Parameter Tuning 793
Heike Trautmann, Boris Naujoks, and Mike Preuss

A Novel Multi-Objective Target Value Optimization Approach 801
S. Wenzel, S. Straatmann, L. Kwiatkowski, P. Schmelzer, and J. Kunert

Desirability-Based Multi-Criteria Optimisation of HVOF Spray Experiments 811
Gerd Kopp, Ingor Baumann, Evelina Vogli, Wolfgang Tillmann, and Claus Weihs

Index 819