Contents

Part I. (Semi-) Plenary Presentations

Classification and Data Mining in Musicology .................. 3
Jan Beran

Bayesian Mixed Membership Models for Soft Clustering and
Classification .................................................... 11
Elena A. Erosheva, Stephen E. Fienberg

Predicting Protein Secondary Structure with Markov Models. 27
Paul Fischer, Simon Larsen, Claus Thomsen

Milestones in the History of Data Visualization: A Case Study
in Statistical Historiography ................................. 34
Michael Friendly

Quantitative Text Typology: The Impact of Word Length ..... 53
Peter Grzybek, Ernst Studtlober, Emmerich Kelih, Gordana Antić

Cluster Ensembles .............................................. 65
Kurt Hornik

Bootstrap Confidence Intervals for Three-way Component
Methods .......................................................... 73
Henk A.L. Kiers

Organising the Knowledge Space for Software Components ... 85
Claus Pahl

Multimedia Pattern Recognition in Soccer Video Using Time
Intervals .......................................................... 97
Cees G.M. Snoek, Marcel Worring

Quantitative Assessment of the Responsibility for the Disease
Load in a Population .......................................... 109
Wolfgang Uter, Olaf Gefeller
Part II. Classification and Data Analysis

Classification

Bootstrapping Latent Class Models 121  
*José G. Dias*

Dimensionality of Random Subspaces 129  
*Eugeniusz Gatnar*

Two-stage Classification with Automatic Feature Selection for an Industrial Application 137  
*Sören Hader, Fred A. Hamprecht*

Bagging, Boosting and Ordinal Classification 145  
*Klaus Hechenbichler, Gerhard Tutz*

A Method for Visual Cluster Validation 153  
*Christian Hennig*

Empirical Comparison of Boosting Algorithms 161  
*Riadh Khanchel, Mohamed Limam*

Iterative Majorization Approach to the Distance-based Discriminant Analysis 168  
*Serhiy Kosinov, Stéphane Marchand-Maillet, Thierry Pun*

An Extension of the CHAID Tree-based Segmentation Algorithm to Multiple Dependent Variables 176  
*Jay Magidson, Jeroen K. Vermunt*

Expectation of Random Sets and the ‘Mean Values’ of Interval Data 184  
*Ole Nordhoff*

Experimental Design for Variable Selection in Data Bases 192  
*Constanze Pumpliün, Claus Weihs, Andrea Preusser*

KMC/EDAM: A New Approach for the Visualization of K-Means Clustering Results 200  
*Nils Raabe, Karsten Luebke, Claus Weihs*
Clustering of Variables with Missing Data: Application to Preference Studies ........................................... 208
  Karin Sahmer, Evelyne Vigneau, El Mostafa Qannari, Joachim Kunert

Binary On-line Classification Based on Temporally Integrated Information ........................................... 216
  Christin Schäfer, Steven Lemm, Gabriel Curio

Different Subspace Classification ........................................... 224
  Gero Szepannek, Karsten Luebke

Density Estimation and Visualization for Data Containing Clusters of Unknown Structure ......................... 232
  Alfred Ultsch

Hierarchical Mixture Models for Nested Data Structures ..... 240
  Jeroen K. Vermunt, Jay Magidson

Data Analysis

Iterative Proportional Scaling Based on a Robust Start Estimator ......................................................... 248
  Claudia Becker

Exploring Multivariate Data Structures with Local Principal Curves ..................................................... 256
  Jochen Einbeck, Gerhard Tutz, Ludger Evers

A Three-way Multidimensional Scaling Approach to the Analysis of Judgments About Persons ..................... 264
  Sabine Krolak-Schwerdt

Discovering Temporal Knowledge in Multivariate Time Series 272
  Fabian Mörchen, Alfred Ultsch

A New Framework for Multidimensional Data Analysis ...... 280
  Shizuhiko Nishisato

External Analysis of Two-mode Three-way Asymmetric Multidimensional Scaling .................................... 288
  Akinori Okada, Tadashi Imaizumi

The Relevance Vector Machine Under Covariate Measurement Error ...................................................... 296
  David Rummel
Part III. Applications

Archaeology

A Contribution to the History of Seriation in Archaeology . . . . . 307
   Peter Ihm

Model-based Cluster Analysis of Roman Bricks and Tiles from
Worms and Rheinzabern ........................................ 317
   Hans-Joachim Mucha, Hans-Georg Bartel, Jens Dolata

Astronomy

Astronomical Object Classification and Parameter Estimation
with the Gaia Galactic Survey Satellite .......................... 325
   Coryn A.L. Bailer-Jones

Design of Astronomical Filter Systems for Stellar
Classification Using Evolutionary Algorithms .................... 330
   Coryn A.L. Bailer-Jones

Bio-Sciences

Analyzing Microarray Data with the Generative Topographic
Mapping Approach ................................................. 338
   Isabelle M. Grimmenstein, Karsten Quast, Wolfgang Urfer

Test for a Change Point in Bernoulli Trials with Dependence . 346
   Joachim Krauth

Data Mining in Protein Binding Cavities .......................... 354
   Katrin Kupas, Alfred Ultsch

Classification of In Vivo Magnetic Resonance Spectra ............ 362
   Björn H. Menze, Michael Wormit, Peter Bachert, Matthias Lichy,
   Heinz-Peter Schlemmer, Fred A. Hamprecht

Modifying Microarray Analysis Methods for Categorical Data
   – SAM and PAM for SNPs ..................................... 370
   Holger Schwender

Improving the Identification of Differentially Expressed Genes
in cDNA Microarray Experiments ............................... 378
   Alfred Ultsch
PhyNav: A Novel Approach to Reconstruct Large Phylogenies .................................................. 386
    Le Sy Vinh, Heiko A. Schmidt, Arndt von Haeseler

Electronic Data and Web

NewsRec, a Personal Recommendation System for News Websites ........................................... 394
    Christian Bomhardt, Wolfgang Gaul

Clustering of Large Document Sets with Restricted Random Walks on Usage Histories .................. 402
    Markus Franke, Anke Theede

Fuzzy Two-mode Clustering vs. Collaborative Filtering ........ 410
    Volker Schlecht, Wolfgang Gaul

Web Mining and Online Visibility ................................. 418
    Nadine Schmidt-Münz, Wolfgang Gaul

Analysis of Recommender System Usage by Multidimensional Scaling ....................................... 426
    Patrick Thoma, Wolfgang Gaul

Finance and Insurance

On a Combination of Convex Risk Minimization Methods .... 434
    Andreas Christmann

Credit Scoring Using Global and Local Statistical Models .... 442
    Alexandra Schwarz, Gerhard Armingier

Informative Patterns for Credit Scoring Using Linear SVM ... 450
    Ralf Stecking, Klaus B. Schebesch

Application of Support Vector Machines in a Life Assurance Environment .................................... 458
    Sarel J. Steel, Gertrud K. Hechter

Continuous Market Risk Budgeting in Financial Institutions.. 466
    Mario Straßberger

Smooth Correlation Estimation with Application to Portfolio Credit Risk .................................... 474
    Rafael Weissbach and Bernd Rosenow
Contents

Libr ary Science and Linguistics

How Many Lexical-semantic Relations are Necessary? .......... 482
  Dariusch Bagheri

Automated Detection of Morphemes Using Distributional
Measur ements ........................................... 490
  Christoph Bendon

Classification of Author and/or Genre? The Impact of Word
Length ...................................................... 498
  Emmerich Keilh, Gordana Antić, Peter Grzybek, Ernst Stadlober

Some Historical Remarks on Library Classification – a Short
Introduction to the Science of Library Classification .......... 506
  Bernd Lorenz

Automatic Validation of Hierarchical Cluster Analysis with
Applic ation in Dialectometry ................................ 513
  Hans-Joachim Mucha, Edgar Haimerl

Discovering the Senses of an Ambiguous Word by Clustering
its Local Contexts ........................................ 521
  Reinhard Rapp

Document Management and the Development of Information
Spaces ...................................................... 529
  Ulfert Rist

Macro-Economics

Stochastic Ranking and the Volatility "Croissant":
A Sensitivity Analysis of Economic Rankings ..................... 537
  Helmut Berrer, Christian Helmenstein, Wolfgang Polasek

Importance Asses sment of Correlated Predictors in Business
Cycles Classification ...................................... 545
  Daniel Enache, Claus Weih

Economic Freedom in the 25-Member European Union:
Insights Using Classification Tools ............................. 553
  Clifford W. Sell

Marketing

Intercultural Consumer Classifications in E-Commerce .......... 561
  Hans H. Bauer, Marcus M. Neumann, Frank Huber
Reservation Price Estimation by Adaptive Conjoint Analysis  . 569
Christoph Breidert, Michael Hahsler, Lars Schmidt-Thieme

Estimating Reservation Prices for Product Bundles Based on Paired Comparison Data .......................... 577
Bernd Stauß, Wolfgang Gaul

Music Science

Classification of Perceived Musical Intervals .................. 585
Jobst P. Fricke

In Search of Variables Distinguishing Low and High Achievers in a Music Sight Reading Task .................. 593
Reinhard Kopiez, Claus Weihs, Uwe Ligges, Ji In Lee

Automatic Feature Extraction from Large Time Series .. 600
Ingo Mierswa

Identification of Musical Instruments by Means of the Hough-Transformation .................................. 608
Christian Röver, Frank Klefenz, Claus Weihs

Support Vector Machines for Bass and Snare Drum Recognition ......................................................... 616
Dirk Van Steelandt, Koen Tanghe, Sven Degroeve, Bernard De Baets, Marc Leman, Jean-Pierre Martens

Register Classification by Timbre ............................... 624
Claus Weihs, Christoph Reuter, Uwe Ligges

Quality Assurance

Classification of Processes by the Lyapunov Exponent ....... 632
Anja M. Busse

Desirability to Characterize Process Capability ............... 640
Jutta Jessenberger, Claus Weihs

Application and Use of Multivariate Control Charts in a BTA Deep Hole Drilling Process ......................... 648
Amor Messaoud, Winfried Theis, Claus Weihs, Franz Hering

Determination of Relevant Frequencies and Modeling Varying Amplitudes of Harmonic Processes ............... 656
Winfried Theis, Claus Weihs
Part IV. Contest: Social Milieus in Dortmund

Introduction to the Contest “Social Milieus in Dortmund” ... 667
Ernst-Otto Sommerer, Claus Weihs

Application of a Genetic Algorithm to Variable Selection in
Fuzzy Clustering .................................................. 674
Christian Röver, Gero Szepannek

Annealed $k$-Means Clustering and Decision Trees .......... 682
Christin Schäfer, Julian Laub

Correspondence Clustering of Dortmund City Districts ...... 690
Stefanie Scheid

Keywords ............................................................... 698

Authors ............................................................. 703
Classification - the Ubiquitous Challenge
Proceedings of the 28th Annual Conference of the
Gesellschaft für Klassifikation e.V., University of
Dortmund, March 9-11, 2004
Weihs, C.; Gaul, W. (Eds.)
2005, XX, 704 p. 181 illus., Softcover
ISBN: 978-3-540-25677-9