

Contents

Plenaries and Semi-plenaries

Boosting and ℓ^1 -Penalty Methods for High-dimensional Data with Some Applications in Genomics	1
<i>P. Bühlmann</i>	
Striving for an Adequate Vocabulary: Next Generation 'Metadata'	13
<i>D. Fellner and S. Havemann</i>	
Scalable Swarm Based Fuzzy Clustering	21
<i>L.O. Hall and P.M. Kanade</i>	
SolEuNet: Selected Data Mining Techniques and Applications	32
<i>N. Lavrač</i>	
Inferred Causation Theory: Time for a Paradigm Shift in Marketing Science?	40
<i>J.A. Mazanec</i>	
Text Mining in Action	52
<i>D. Mladenič</i>	
Identification of Real-world Objects in Multiple Databases	63
<i>M. Neiling</i>	
Kernels for Predictive Graph Mining	75
<i>S. Wrobel, T. Gärtner, and T. Horváth</i>	

Clustering

PRISMA: Improving Risk Estimation with Parallel Logistic Regression Trees	87
<i>B. Arrnrich, A. Albert, and J. Walter</i>	
Latent Class Analysis and Model Selection	95
<i>J.G. Dias</i>	

An Indicator for the Number of Clusters:
Using a Linear Map to Simplex Structure 103
M. Weber, W. Rungtarityotin, and A. Schliep

Discriminant Analysis

On the Use of Some Classification Quality Measure
to Construct Mean Value Estimates Under Nonresponse 111
W. Gamrot

A Wrapper Feature Selection Method
for Combined Tree-based Classifiers 119
E. Gatnar

Input Variable Selection
in Kernel Fisher Discriminant Analysis 126
N. Louw and S.J. Steel

The Wavelet Packet Based Cepstral Features
for Open Set Speaker Classification in Marathi 134
H.A. Patil, P.K. Dutta, and T.K. Basu

A New Effective Algorithm for Stepwise
Principle Components Selection in Discriminant Analysis 142
E. Serikova and E. Zhuk

A Comparison of Validation Methods
for Learning Vector Quantization and for Support Vector Machines
on Two Biomedical Data Sets 150
D. Sommer and M. Golz

Discriminant Analysis of Polythetically Described
Older Palaeolithic Stone Flakes: Possibilities and Questions 158
T. Weber

Classification with Latent Variable Models

Model-based Density Estimation by Independent Factor Analysis 166
D.G. Calò, A. Montanari, and C. Viroli

Identifying Multiple Cluster Structures
Through Latent Class Models 174
G. Galimberti and G. Soffritti

Gene Selection in Classification Problems via Projections onto a Latent Space	182
<i>M. Pillati and C. Viroli</i>	

Multiway Classification and Data Analysis

The Recovery Performance of Two-mode Clustering Methods: Monte Carlo Experiment	190
<i>S. Krolak-Schwerdt and M. Wiedenbeck</i>	
On the Comparability of Reliability Measures: Bifurcation Analysis of Two Measures in the Case of Dichotomous Ratings	198
<i>T. Ostermann and R. Schuster</i>	

Ranking, Multi-label Classification, Preferences

On Active Learning in Multi-label Classification	206
<i>K. Brinker</i>	
From Ranking to Classification: A Statistical View	214
<i>S. Cl��men��on, G. Lugosi, and N. Vayatis</i>	

PLS Path Modeling, PLS Regression and Classification

Assessing Unidimensionality within PLS Path Modeling Framework ...	222
<i>K. Sahmer, M. Hanafi, and E.M. Qannari</i>	
The Partial Robust M-approach	230
<i>S. Serneels, C. Croux, P. Filzmoser, and P.J. Van Espen</i>	
Classification in PLS Path Models and Local Model Optimisation	238
<i>S. Squillacciotti</i>	

Robust Methods in Multivariate Statistics

Hierarchical Clustering by Means of Model Grouping	246
<i>C. Agostinelli and P. Pellizzari</i>	
Deepest Points and Least Deep Points: Robustness and Outliers with MZE	254
<i>C. Becker and S.P. Scholz</i>	

Robust Transformations and Outlier Detection with Autocorrelated Data	262
<i>A. Cerioli and M. Riani</i>	
Robust Multivariate Methods: The Projection Pursuit Approach	270
<i>P. Filzmoser, S. Serneels, C. Croux, and P.J. Van Espen</i>	
Finding Persisting States for Knowledge Discovery in Time Series	278
<i>F. Mörchen and A. Ultsch</i>	

Data Mining and Explorative Multivariate Data Analysis

Restricted Co-inertia Analysis	286
<i>P. Amenta and E. Ciavolino</i>	
Hausman Principal Component Analysis	294
<i>V. Choulakian, L. Dambra, and B. Simonetti</i>	
Nonlinear Time Series Modelling: Monitoring a Drilling Process	302
<i>A. Messaoud, C. Weihs, and F. Hering</i>	

Text Mining

Word Length and Frequency Distributions in Different Text Genres ...	310
<i>G. Antić, E. Stadlober, P. Grzybek, and E. Kelih</i>	
Bootstrapping an Unsupervised Morphemic Analysis	318
<i>C. Benden</i>	
Automatic Extension of Feature-based Semantic Lexicons via Contextual Attributes	326
<i>C. Biemann and R. Osswald</i>	
Learning Ontologies to Improve Text Clustering and Classification	334
<i>S. Bloehdorn, P. Cimiano, and A. Hotho</i>	
Discovering Communities in Linked Data by Multi-view Clustering	342
<i>I. Drost, S. Bickel, and T. Scheffer</i>	
Crosslinguistic Computation and a Rhythm-based Classification of Languages	350
<i>A. Fenk and G. Fenk-Oczlon</i>	
Using String Kernels for Classification of Slovenian Web Documents ...	358
<i>B. Fortuna and D. Mladenić</i>	

Semantic Decomposition of Character Encodings for Linguistic Knowledge Discovery	366
<i>D. Gibbon, B. Hughes, and T. Trippel</i>	
Applying Collaborative Filtering to Real-life Corporate Data	374
<i>M. Grcar, D. Mladenič, and M. Grobelnik</i>	
Quantitative Text Typology: The Impact of Sentence Length	382
<i>E. Kelih, P. Grzybek, G. Antić, and E. Stadlober</i>	
A Hybrid Machine Learning Approach for Information Extraction from Free Text	390
<i>G. Neumann</i>	
Text Classification with Active Learning	398
<i>B. Novak, D. Mladenič, and M. Grobelnik</i>	
Towards Structure-sensitive Hypertext Categorization	406
<i>A. Mehler, R. Gleim, and M. Dehmer</i>	
Evaluating the Performance of Text Mining Systems on Real-world Press Archives	414
<i>G. Paaß and H. de Vries</i>	
Part-of-Speech Induction by Singular Value Decomposition and Hierarchical Clustering	422
<i>R. Rapp</i>	
Near Similarity Search and Plagiarism Analysis	430
<i>B. Stein and S.M. zu Eissen</i>	

Fuzzy Data Analysis

Objective Function-based Discretization	438
<i>F. Höppner</i>	
Understanding and Controlling the Membership Degrees in Fuzzy Clustering	446
<i>F. Klawonn</i>	
Autonomous Sensor-based Landing Systems: Fusion of Vague and Incomplete Information by Application of Fuzzy Clustering Techniques	454
<i>B. Korn</i>	
Outlier Preserving Clustering for Structured Data Through Kernels . . .	462
<i>M.-J. Lesot</i>	

Economics and Mining in Business Processes

Classification-relevant Importance Measures for the West German Business Cycle	470
<i>D. Enache, C. Weihs, and U. Garczarek</i>	
The Classification of Local and Branch Labour Markets in the Upper Silesia	478
<i>W. Hantke</i>	
An Overview of Artificial Life Approaches for Clustering	486
<i>D. Kämpf and A. Ultsch</i>	
Design Problems of Complex Economic Experiments	494
<i>J. Kunze</i>	
Traffic Sensitivity of Long-term Regional Growth Forecasts	502
<i>W. Polasek and H. Berrer</i>	
Spiralling in BTA Deep-hole Drilling: Models of Varying Frequencies	510
<i>N. Raabe, O. Webber, W. Theis, and C. Weihs</i>	
Analysis of the Economic Development of Districts in Poland as a Basis for the Framing of Regional Policies	518
<i>M. Rozkrut and D. Rozkrut</i>	

Banking and Finance

The Classification of Candlestick Charts: Laying the Foundation for Further Empirical Research	526
<i>S. Etschberger, H. Fock, C. Klein, and B. Zwergel</i>	
Modeling and Estimating the Credit Cycle by a Probit-AR(1)-Process	534
<i>S. Höse and K. Vogl</i>	
Comparing and Selecting SVM-Kernels for Credit Scoring	542
<i>R. Stecking and K.B. Schebesch</i>	
Value at Risk Using the Principal Components Analysis on the Polish Power Exchange	550
<i>G. Trzpiot and A. Ganczarek</i>	

Marketing

- A Market Basket Analysis
Conducted with a Multivariate Logit Model 558
Y. Boztuğ and L. Hildebrandt
- Solving and Interpreting Binary Classification Problems in Marketing
with SVMs 566
G. Nalbantov, J.C. Bioch, and P.J.F. Groenen
- Modeling the Nonlinear Relationship Between Satisfaction and Loyalty
with Structural Equation Models 574
M. Paulssen and A. Sommerfeld
- Job Choice Model to Measure Behavior
in a Multi-stage Decision Process 582
T. Spengler and J. Malmendier
- Semiparametric Stepwise Regression
to Estimate Sales Promotion Effects 590
W.J. Steiner, C. Belätz, and S. Lang

Adaptivity and Personalization

- Implications of Probabilistic Data Modeling
for Mining Association Rules 598
M. Hahsler, K. Hornik, and T. Reutterer
- Copula Functions in Model Based Clustering 606
K. Jajuga and D. Papla
- Attribute-aware Collaborative Filtering 614
K. Tso and L. Schmidt-Thieme

User and Data Authentication in IT Security

- Towards a Flexible Framework for Open Source Software
for Handwritten Signature Analysis 622
R. Guest, M. Fairhurst, and C. Vielhauer
- Multimodal Biometric Authentication System
Based on Hand Features 630
N. Pavešić, T. Savič, and S. Ribarić

Labelling and Authentication for Medical Imaging Through Data Hiding	638
<i>A. De Rosa, R. Caldelli, and A. Piva</i>	

Hand-geometry Recognition Based on Contour Landmarks	646
<i>R. Veldhuis, A. Bazen, W. Booij, and A. Hendrikse</i>	

A Cross-cultural Evaluation Framework for Behavioral Biometric User Authentication.....	654
<i>F. Wolf, T.K. Basu, P.K. Dutta, C. Vielhauer, A. Oermann, and B. Yegnanarayana</i>	

Bioinformatics and Biostatistics

On External Indices for Mixtures: Validating Mixtures of Genes	662
<i>I.G. Costa and A. Schliep</i>	

Tests for Multiple Change Points in Binary Markov Sequences	670
<i>J. Krauth</i>	

UnitExpressions: A Rational Normalization Scheme for DNA Microarray Data	678
<i>A. Ultsch</i>	

Classification of High-dimensional Biological and Medical Data

A Ridge Classification Method for High-dimensional Observations	684
<i>M. Grüning and S. Kropf</i>	

Assessing the Trustworthiness of Clustering Solutions Obtained by a Function Optimization Scheme	692
<i>U. Möller and D. Radke</i>	

Variable Selection for Discrimination of More Than Two Classes Where Data are Sparse	700
<i>G. Szepannek and C. Weihs</i>	

Medical and Health Sciences

The Assessment of Second Primary Cancers (SPCs) in a Series of Splenic Marginal Zone Lymphoma (SMZL) Patients	708
<i>S. De Cantis and A.M. Taormina</i>	

Heart Rate Classification Using Support Vector Machines	716
<i>M. Vogt, U. Moissl, and J. Schaab</i>	

Music Analysis

Visual Mining in Music Collections	724
<i>F. Mörchén, A. Ultsch, M. Nöcker, and C. Stamm</i>	
Modeling Memory for Melodies	732
<i>D. Müllensiefen and C. Hennig</i>	
Parameter Optimization in Automatic Transcription of Music	740
<i>C. Weihs and U. Ligges</i>	

Data Mining Competition

GfKl Data Mining Competition 2005:	
Predicting Liquidity Crises of Companies	748
<i>J. Strackeljan, R. Jonscher, S. Prieur, D. Vogel, T. Deselaers,</i>	
<i>D. Keysers, A. Mauser, I. Bezrukov, and A. Hegerath</i>	
Author Index	759