

Contents

Part I Clustering

Mixture Models for Classification

Gilles Celeux 3

How to Choose the Number of Clusters: The Cramer Multiplicity Solution

Adriana Climescu-Haulica 15

Model Selection Criteria for Model-Based Clustering of Categorical Time Series Data: A Monte Carlo Study

José G. Dias 23

Cluster Quality Indexes for Symbolic Classification – An Examination

Andrzej Dudek 31

Semi-Supervised Clustering: Application to Image Segmentation

Mário A.T. Figueiredo 39

A Method for Analyzing the Asymptotic Behavior of the Walk Process in Restricted Random Walk Cluster Algorithm

Markus Franke, Andreas Geyer-Schulz 51

Cluster and Select Approach to Classifier Fusion

Eugeniusz Gatnar 59

Random Intersection Graphs and Classification

Erhard Godehardt, Jerzy Jaworski, Katarzyna Rybarczyk 67

Optimized Alignment and Visualization of Clustering Results

Martin Hoffmann, Dörte Radke, Ulrich Möller 75

Finding Cliques in Directed Weighted Graphs Using Complex Hermitian Adjacency Matrices	
<i>Bettina Hoser, Thomas Bierhance</i>	83
Text Clustering with String Kernels in R	
<i>Alexandros Karatzoglou, Ingo Feinerer</i>	91
Automatic Classification of Functional Data with Extremal Information	
<i>Fabrizio Laurini, Andrea Cerioli</i>	99
Typicality Degrees and Fuzzy Prototypes for Clustering	
<i>Marie-Jeanne Lesot, Rudolf Kruse</i>	107
On Validation of Hierarchical Clustering	
<i>Hans-Joachim Mucha</i>	115
<hr/>	
Part II Classification	
<hr/>	
Rearranging Classified Items in Hierarchies Using Categorization Uncertainty	
<i>Korinna Bade, Andreas Nürnberger</i>	125
Localized Linear Discriminant Analysis	
<i>Irina Czogiel, Karsten Luebke, Marc Zentgraf, Claus Weihs</i>	133
Calibrating Classifier Scores into Probabilities	
<i>Martin Gebel, Claus Weihs</i>	141
Nonlinear Support Vector Machines Through Iterative Majorization and I-Splines	
<i>Patrick J.F. Groenen, Georgi Nalbantov, J. Cor Bioch</i>	149
Deriving Consensus Rankings from Benchmarking Experiments	
<i>Kurt Hornik, David Meyer</i>	163
Classification of Contradiction Patterns	
<i>Heiko Müller, Ulf Leser, Johann-Christoph Freytag</i>	171
Selecting SVM Kernels and Input Variable Subsets in Credit Scoring Models	
<i>Klaus B. Schebesch, Ralf Stecking</i>	179

Part III Data and Time Series Analysis

**Simultaneous Selection of Variables and Smoothing
Parameters in Geoadditive Regression Models**
Christiane Belitz, Stefan Lang 189

Modelling and Analysing Interval Data
Paula Brito 197

**Testing for Genuine Multimodality in Finite Mixture Models:
Application to Linear Regression Models**
Bettina Grün, Friedrich Leisch 209

Happy Birthday to You, Mr. Wilcoxon!
Invariance, Semiparametric Efficiency, and Ranks
Marc Hallin 217

**Equivalent Number of Degrees of Freedom for Neural
Networks**
Salvatore Ingrassia, Isabella Morlini 229

**Model Choice for Panel Spatial Models: Crime Modeling in
Japan**
Kazuhiko Kakamu, Wolfgang Polasek, Hajime Wago 237

A Boosting Approach to Generalized Monotonic Regression
Florian Leitenstorfer, Gerhard Tutz 245

**From Eigenspots to Fisherspots – Latent Spaces in the
Nonlinear Detection of Spot Patterns in a Highly Varying
Background**
Bjoern H. Menze, B. Michael Kelm, Fred A. Hamprecht 255

Identifying and Exploiting Ultrametricity
Fionn Murtagh 263

**Factor Analysis for Extraction of Structural Components and
Prediction in Time Series**
Carsten Schneider, Gerhard Arminger 273

**Classification of the U.S. Business Cycle by Dynamic Linear
Discriminant Analysis**
Roland Schuhr 281

Examination of Several Results of Different Cluster Analyses with a Separate View to Balancing the Economic and Ecological Performance Potential of Towns and Cities	
<i>Nguyen Xuan Thinh, Martin Behnisch, Alfred Ultsch</i>	289

Part IV Visualization and Scaling Methods

VOS: A New Method for Visualizing Similarities Between Objects	
<i>Nees Jan van Eck, Ludo Waltman</i>	299
Multidimensional Scaling of Asymmetric Proximities with a Dominance Point	
<i>Akinori Okada, Tadashi Imaizumi</i>	307
Single Cluster Visualization to Optimize Air Traffic Management	
<i>Frank Rehm, Frank Klawonn, Rudolf Kruse</i>	319
Rescaling Proximity Matrix Using Entropy Analyzed by INDSCAL	
<i>Satoru Yokoyama, Akinori Okada</i>	327

Part V Information Retrieval, Data and Web Mining

Canonical Forms for Frequent Graph Mining	
<i>Christian Borgelt</i>	337
Applying Clickstream Data Mining to Real-Time Web Crawler Detection and Containment Using ClickTips Platform	
<i>Anália Lourenço, Orlando Belo</i>	351
Plagiarism Detection Without Reference Collections	
<i>Sven Meyer zu Eissen, Benno Stein, Marion Kulig</i>	359
Putting Successor Variety Stemming to Work	
<i>Benno Stein, Martin Potthast</i>	367
Collaborative Filtering Based on User Trends	
<i>Panagiotis Symeonidis, Alexandros Nanopoulos, Apostolos Papadopoulos, Yannis Manolopoulos</i>	375
Investigating Unstructured Texts with Latent Semantic Analysis	
<i>Fridolin Wild, Christina Stahl</i>	383

Part VI Marketing, Management Science and Economics

Heterogeneity in Preferences for Odd Prices*Bernhard Baumgartner, Winfried J. Steiner* 393**Classification of Reference Models***Robert Braun, Werner Esswein* 401**Adaptive Conjoint Analysis for Pricing Music Downloads***Christoph Breidert, Michael Hahsler* 409**Improving the Probabilistic Modeling of Market Basket Data***Christian Buchta*..... 417**Classification in Marketing Research by Means of
LEM2-generated Rules***Reinhold Decker, Frank Kroll*..... 425**Pricing Energy in a Multi-Utility Market***Markus Franke, Andreas Kamper, Anke Eßer* 433**Disproportionate Samples in Hierarchical Bayes CBC
Analysis***Sebastian Fuchs, Manfred Schwaiger* 441**Building on the Arules Infrastructure for Analyzing
Transaction Data with R***Michael Hahsler, Kurt Hornik* 449**Balanced Scorecard Simulator – A Tool for Stochastic
Business Figures***Veit Köppen, Marina Allgeier, Hans-J. Lenz*..... 457**Integration of Customer Value into Revenue Management***Tobias von Martens, Andreas Hilbert* 465**Women's Occupational Mobility and Segregation in the
Labour Market: Asymmetric Multidimensional Scaling***Miki Nakai* 473**Multilevel Dimensions of Consumer Relationships in the
Healthcare Service Market M-L IRT vs. M-L SEM Approach***Iga Rudawska, Adam Sagan* 481

Data Mining in Higher Education

Karoline Schönbrunn, Andreas Hilbert 489

Attribute Aware Anonymous Recommender Systems

Manuel Stritt, Karen H.L. Tso, Lars Schmidt-Thieme 497

Part VII Banking and Finance

On the Notions and Properties of Risk and Risk Aversion in the Time Optimal Approach to Decision Making

Martin Bouzaima, Thomas Burkhardt 507

A Model of Rational Choice Among Distributions of Goal Reaching Times

Thomas Burkhardt 515

On Goal Reaching Time Distributions Estimated from DAX Stock Index Investments

Thomas Burkhardt, Michael Haasis 523

Credit Risk of Collaterals: Examining the Systematic Linkage between Insolvencies and Physical Assets in Germany

Marc Gürtler, Dirk Heithecker, Sven Olboeter 531

Foreign Exchange Trading with Support Vector Machines

Christian Ullrich, Detlef Seese, Stephan Chalup 539

The Influence of Specific Information on the Credit Risk Level

Miroslaw Wójciak, Aleksandra Wójcicka-Krenz 547

Part VIII Bio- and Health Sciences

Enhancing Bluejay with Scalability, Genome Comparison and Microarray Visualization

Anguo Dong, Andrei L. Turinsky, Andrew C. Ah-Seng, Morgan Taschuk, Paul M.K. Gordon, Katharina Hochauer, Sabrina Fröls, Jung Soh, Christoph W. Sensen 557

Discovering Biomarkers for Myocardial Infarction from SELDI-TOF Spectra

Christian Höner zu Siederdissen, Susanne Ragg, Sven Rahmann 569

Joint Analysis of In-situ Hybridization and Gene Expression Data

Lennart Opitz, Alexander Schliep, Stefan Posch 577

Unsupervised Decision Trees Structured by Gene Ontology (GO-UDTs) for the Interpretation of Microarray Data <i>Henning Redestig, Florian Sohler, Ralf Zimmer, Joachim Selbig</i>	585
--	-----

Part IX Linguistics and Text Analysis

Clustering of Polysemic Words <i>Laurent Cicurel, Stephan Bloehdorn, Philipp Cimiano</i>	595
--	-----

Classifying German Questions According to Ontology-Based Answer Types <i>Adriana Davidescu, Andrea Heyl, Stefan Kazalski, Irene Cramer, Dietrich Klakow</i>	603
---	-----

The Relationship of Word Length and Sentence Length: The Inter-Textual Perspective <i>Peter Grzybek, Ernst Stadlober, Emmerich Kelih</i>	611
--	-----

Comparing the Stability of Different Clustering Results of Dialect Data <i>Edgar Haimperl, Hans-Joachim Mucha</i>	619
---	-----

Part-of-Speech Discovery by Clustering Contextual Features <i>Reinhard Rapp</i>	627
---	-----

Part X Statistical Musicology and Sound Classification

A Probabilistic Framework for Audio-Based Tonal Key and Chord Recognition <i>Benoit Catteau, Jean-Pierre Martens, Marc Leman</i>	637
--	-----

Using MCMC as a Stochastic Optimization Procedure for Monophonic and Polyphonic Sound <i>Katrin Sommer, Claus Weihs</i>	645
---	-----

Vowel Classification by a Neurophysiologically Parameterized Auditory Model <i>Gero Szepannek, Tamás Harczos, Frank Klefenz, András Katai, Patrick Schikowski, Claus Weihs</i>	653
--	-----

Part XI Archaeology

**Uncovering the Internal Structure of the Roman Brick and
Tile Making in Frankfurt-Nied by Cluster Validation**
Jens Dolata, Hans-Joachim Mucha, Hans-Georg Bartel 663

**Where Did I See You Before...
A Holistic Method to Compare and Find Archaeological
Artifacts**
Vincent Mom..... 671

Keywords..... 681

Author Index..... 685

Advances in Data Analysis

Proceedings of the 30th Annual Conference of the
Gesellschaft für Klassifikation e.V., Freie Universität
Berlin, March 8-10, 2006

Decker, R.; Lenz, H.-J. (Eds.)

2007, XVI, 687 p., Softcover

ISBN: 978-3-540-70980-0