

Contents

Part I Classification

Distance-based Kernels for Real-valued Data

Lluís Belanche, Jean Luis Vázquez, Miguel Vázquez 3

Fast Support Vector Machine Classification of Very Large Datasets

Janis Fehr, Karina Zapién Arreola, Hans Burkhardt 11

Fusion of Multiple Statistical Classifiers

Eugeniusz Gatnar 19

Calibrating Margin-based Classifier Scores into Polychotomous Probabilities

Martin Gebel, Claus Weihs 29

Classification with Invariant Distance Substitution Kernels

Bernard Haasdonk, Hans Burkhardt 37

Applying the Kohonen Self-organizing Map Networks to Select Variables

Kamila Migdał Najman, Krzysztof Najman 45

Computer Assisted Classification of Brain Tumors

Norbert Röhl, José R. Iglesias-Rozas, Galia Weidl 55

Model Selection in Mixture Regression Analysis – A Monte Carlo Simulation Study

Marko Sarstedt, Manfred Schwaiger 61

Comparison of Local Classification Methods

Julia Schiffner, Claus Weihs 69

Incorporating Domain Specific Information into Gaia Source Classification

Kester W. Smith, Carola Tiede, Coryn A.L. Bailer-Jones 77

Identification of Noisy Variables for Nonmetric and Symbolic Data in Cluster Analysis	
<i>Marek Walesiak, Andrzej Dudek</i>	85

Part II Clustering

Families of Dendrograms	
<i>Patrick Erik Bradley</i>	95
Mixture Models in Forward Search Methods for Outlier Detection	
<i>Daniela G. Calò</i>	103
On Multiple Imputation Through Finite Gaussian Mixture Models	
<i>Marco Di Zio, Ugo Guarnera</i>	111
Mixture Model Based Group Inference in Fused Genotype and Phenotype Data	
<i>Benjamin Georgi, M.Anne Spence, Pamela Flodman, Alexander Schliep</i>	119
The Noise Component in Model-based Cluster Analysis	
<i>Christian Hennig, Pietro Coretto</i>	127
An Artificial Life Approach for Semi-supervised Learning	
<i>Lutz Herrmann, Alfred Ultsch</i>	139
Hard and Soft Euclidean Consensus Partitions	
<i>Kurt Hornik, Walter Böhm</i>	147
Rationale Models for Conceptual Modeling	
<i>Sina Lehrmann, Werner Esswein</i>	155
Measures of Dispersion and Cluster-Trees for Categorical Data	
<i>Ulrich Müller-Funk</i>	163
Information Integration of Partially Labeled Data	
<i>Steffen Rendle, Lars Schmidt-Thieme</i>	171

Part III Multidimensional Data Analysis

Data Mining of an On-line Survey - A Market Research Application

Karmele Fernández-Aguirre, María I. Landaluce, Ana Martín, Juan I. Modroño 183

Nonlinear Constrained Principal Component Analysis in the Quality Control Framework

Michele Gallo, Luigi D'Ambra 193

Non Parametric Control Chart by Multivariate Additive Partial Least Squares via Spline

Rosaria Lombardo, Amalia Vanacore, Jean-François Durand 201

Simple Non Symmetrical Correspondence Analysis

Antonello D'Ambra, Pietro Amenta, Valentin Rousson 209

Factorial Analysis of a Set of Contingency Tables

Amaya Zárraga, Beatriz Goitisoló 219

Part IV Analysis of Complex Data

Graph Mining: Repository vs. Canonical Form

Christian Borgelt and Mathias Fiedler 229

Classification and Retrieval of Ancient Watermarks

Gerd Brunner, Hans Burkhardt 237

Segmentation and Classification of Hyper-Spectral Skin Data

Hannes Kazianka, Raimund Leitner, Jürgen Pilz 245

FSMTree: An Efficient Algorithm for Mining Frequent Temporal Patterns

Steffen Kempe, Jochen Hipp, Rudolf Kruse 253

A Matlab Toolbox for Music Information Retrieval

Olivier Lartillot, Petri Toivainen, Tuomas Eerola 261

A Probabilistic Relational Model for Characterizing Situations in Dynamic Multi-Agent Systems

Daniel Meyer-Delius, Christian Plagemann, Georg von Wichert, Wendelin Feiten, Gisbert Lawitzky, Wolfram Burgard 269

Applying the Q_n Estimator Online

Robin Nunkesser, Karen Schettlinger, Roland Fried 277

A Comparative Study on Polyphonic Musical Time Series Using MCMC Methods	
<i>Katrin Sommer, Claus Weihs</i>	285
Collective Classification for Labeling of Places and Objects in 2D and 3D Range Data	
<i>Rudolph Triebel, Óscar Martínez Mozos, Wolfram Burgard</i>	293
Lag or Error? - Detecting the Nature of Spatial Correlation	
<i>Mario Larch, Janette Walde</i>	301
<hr/>	
Part V Exploratory Data Analysis and Tools for Data Analysis	
<hr/>	
Urban Data Mining Using Emergent SOM	
<i>Martin Behnisch, Alfred Ultsch</i>	311
KNIME: The Konstanz Information Miner	
<i>Michael R. Berthold, Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias Kötter, Thorsten Meinl, Peter Ohl, Christoph Sieb, Kilian Thiel, Bernd Wiswedel</i>	319
A Pattern Based Data Mining Approach	
<i>Boris Delibašić, Kathrin Kirchner, Johannes Ruhland</i>	327
A Framework for Statistical Entity Identification in R	
<i>Michaela Denk</i>	335
Combining Several SOM Approaches in Data Mining: Application to ADSL Customer Behaviours Analysis	
<i>Francoise Fessant, Vincent Lemaire, Fabrice Clérot</i>	343
On the Analysis of Irregular Stock Market Trading Behavior	
<i>Markus Franke, Bettina Hoser, Jan Schröder</i>	355
A Procedure to Estimate Relations in a Balanced Scorecard	
<i>Veit Köppen, Henner Graubitz, Hans-K. Arndt, Hans-J. Lenz</i>	363
The Application of Taxonomies in the Context of Configurative Reference Modelling	
<i>Ralf Knackstedt, Armin Stein</i>	373
Two-Dimensional Centrality of a Social Network	
<i>Akinori Okada</i>	381
Benchmarking Open-Source Tree Learners in R/RWeka	
<i>Michael Schauerhuber, Achim Zeileis, David Meyer, Kurt Hornik</i>	389

From Spelling Correction to Text Cleaning – Using Context Information	
<i>Martin Schierle, Sascha Schulz, Markus Ackermann</i>	397
Root Cause Analysis for Quality Management	
<i>Christian Manuel Strobel, Tomas Hrycej</i>	405
Finding New Technological Ideas and Inventions with Text Mining and Technique Philosophy	
<i>Dirk Thorleuchter</i>	413
Investigating Classifier Learning Behavior with Experiment Databases	
<i>Joaquin Vanschoren, Hendrik Blockeel</i>	421
<hr/>	
Part VI Marketing and Management Science	
<hr/>	
Conjoint Analysis for Complex Services Using Clusterwise Hierarchical Bayes Procedures	
<i>Michael Brusch, Daniel Baier</i>	431
Building an Association Rules Framework for Target Marketing	
<i>Nicolas March, Thomas Reutterer</i>	439
AHP versus ACA – An Empirical Comparison	
<i>Martin Meißner, Sören W. Scholz, Reinhold Decker</i>	447
On the Properties of the Rank Based Multivariate Exponentially Weighted Moving Average Control Charts	
<i>Amor Messaoud, Claus Weihs</i>	455
Are Critical Incidents Really Critical for a Customer Relationship? A MIMIC Approach	
<i>Marcel Paulssen, Angela Sommerfeld</i>	463
Heterogeneity in the Satisfaction-Retention Relationship – A Finite-mixture Approach	
<i>Dorian Quint, Marcel Paulssen</i>	471
An Early-Warning System to Support Activities in the Management of Customer Equity and How to Obtain the Most from Spatial Customer Equity Potentials	
<i>Klaus Thiel, Daniel Probst</i>	479
Classifying Contemporary Marketing Practices	
<i>Ralf Wagner</i>	489

Part VII Banking and Finance

- Predicting Stock Returns with Bayesian Vector Autoregressive Models**
Wolfgang Bessler, Peter Lückoff 499
- The Evaluation of Venture-Backed IPOs – Certification Model versus Adverse Selection Model, Which Does Fit Better?**
Francesco Gangi, Rosaria Lombardo 507
- Using Multiple SVM Models for Unbalanced Credit Scoring Data Sets**
Klaus B. Schebesch, Ralf Stecking 515

Part VIII Business Intelligence

- Comparison of Recommender System Algorithms Focusing on the New-item and User-bias Problem**
Stefan Hauger, Karen H. L. Tso, Lars Schmidt-Thieme 525
- Collaborative Tag Recommendations**
Leandro Balby Marinho and Lars Schmidt-Thieme 533
- Applying Small Sample Test Statistics for Behavior-based Recommendations**
Andreas W. Neumann, Andreas Geyer-Schulz 541

Part IX Text Mining, Web Mining, and the Semantic Web

- Classifying Number Expressions in German Corpora**
Irene Cramer, Stefan Schacht, Andreas Merkel 553
- Non-Profit Web Portals - Usage Based Benchmarking for Success Evaluation**
Daniel Delić, Hans-J. Lenz 561
- Text Mining of Supreme Administrative Court Jurisdictions**
Ingo Feinerer, Kurt Hornik 569
- Supporting Web-based Address Extraction with Unsupervised Tagging**
Berenike Loos, Chris Biemann 577
- A Two-Stage Approach for Context-Dependent Hypernym Extraction**
Berenike Loos, Mario DiMarzo 585
- Analysis of Dwell Times in Web Usage Mining**
Patrick Mair, Marcus Hudec 593

New Issues in Near-duplicate Detection	
<i>Martin Potthast, Benno Stein</i>	601
Comparing the University of South Florida Homograph Norms with Empirical Corpus Data	
<i>Reinhard Rapp</i>	611
Content-based Dimensionality Reduction for Recommender Systems	
<i>Panagiotis Symeonidis</i>	619
<hr/>	
Part X Linguistics	
<hr/>	
The Distribution of Data in Word Lists and its Impact on the Subgrouping of Languages	
<i>Hans J. Holm</i>	629
Quantitative Text Analysis Using L-, F- and T-Segments	
<i>Reinhard Köhler, Sven Naumann</i>	637
Projecting Dialect Distances to Geography: Bootstrap Clustering vs. Noisy Clustering	
<i>John Nerbonne, Peter Kleiweg, Wilbert Heeringa, Franz Manni</i>	647
Structural Differentiae of Text Types – A Quantitative Model	
<i>Olga Pustyl'nikov, Alexander Mehler</i>	655
<hr/>	
Part XI Data Analysis in Humanities	
<hr/>	
Scenario Evaluation Using Two-mode Clustering Approaches in Higher Education	
<i>Matthias J. Kaiser, Daniel Baier</i>	665
Visualization and Clustering of Tagged Music Data	
<i>Pascal Lehwark, Sebastian Risi, Alfred Ultsch</i>	673
Effects of Data Transformation on Cluster Analysis of Archaeometric Data	
<i>Hans-Joachim Mucha, Hans-Georg Bartel, Jens Dolata</i>	681
Fuzzy PLS Path Modeling: A New Tool For Handling Sensory Data	
<i>Francesco Palumbo, Rosaria Romano, Vincenzo Esposito Vinzi</i>	689
Automatic Analysis of Dewey Decimal Classification Notations	
<i>Ulrike Reiner</i>	697

XVI Contents

A New Interval Data Distance Based on the Wasserstein Metric <i>Rosanna Verde, Antonio Irpino</i>	705
Keywords	713
Author Index	717

Data Analysis, Machine Learning and Applications
Proceedings of the 31st Annual Conference of the
Gesellschaft für Klassifikation e.V.,
Albert-Ludwigs-Universität Freiburg, March 7-9, 2007
Preisach, C.; Burkhardt, H.; Schmidt-Thieme, B.; Decker,
R. (Eds.)
2008, XVI, 719 p.,
ISBN: 978-3-540-78246-9