

Andreas Fink · Berthold Lausen
Wilfried Seidel · Alfred Ultsch
Editors

Advances in Data Analysis, Data Handling and Business Intelligence

Proceedings of the 32nd Annual Conference
of the Gesellschaft für Klassifikation e.V.,
Joint Conference with the
British Classification Society (BCS) and the
Dutch/Flemish Classification Society (VOC),
Helmut-Schmidt-University, Hamburg,
July 16–18, 2008

 Springer

Contents

Part I Invited

Semi-supervised Probabilistic Distance Clustering and the Uncertainty of Classification	3
Cem Iyigun and Adi Ben-Israel	
Strategies of Model Construction for the Analysis of Judgment Data	21
Sabine Krolak-Schwerdt	
Clustering of High-Dimensional Data via Finite Mixture Models	33
Geoff J. McLachlan and Jangsun Baek	
Clustering and Dimensionality Reduction to Discover Interesting Patterns in Binary Data	45
Francesco Palumbo and Alfonso Iodice D'Enza	
Kernel Methods for Detecting the Direction of Time Series	57
Jonas Peters, Dominik Janzing, Arthur Gretton, and Bernhard Schölkopf	
Statistical Processes Under Change: Enhancing Data Quality with Pretests	67
Walter Radermacher and Sabine Sattelberger	
Part II Clustering and Classification	
Evaluation Strategies for Learning Algorithms of Hierarchies	83
Korinna Bade and Dominik Benz	

Fuzzy Subspace Clustering	93
Christian Borgelt	
Motif-Based Classification of Time Series with Bayesian Networks and SVMs	105
Krisztian Buza and Lars Schmidt-Thieme	
A Novel Approach to Construct Discrete Support Vector Machine Classifiers	115
Marco Caserta, Stefan Lessmann, and Stefan Voß	
Predictive Classification Trees	127
Stephan Dlugosz and Ulrich Müller-Funk	
Isolated Vertices in Random Intersection Graphs	135
Erhard Godehardt, Jerzy Jaworski, and Katarzyna Rybarczyk	
Strengths and Weaknesses of Ant Colony Clustering	147
Lutz Herrmann and Alfred Ultsch	
Variable Selection for Kernel Classifiers: A Feature-to-Input Space Approach	157
Surette Oosthuizen and Sarel Steel	
Finite Mixture and Genetic Algorithm Segmentation in Partial Least Squares Path Modeling: Identification of Multiple Segments in Complex Path Models	167
Christian M. Ringle, Marko Sarstedt, and Rainer Schlittgen	
Cluster Ensemble Based on Co-occurrence Data	177
Dorota Rozmus	
Localized Logistic Regression for Categorical Influential Factors	185
Julia Schiffner, Gero Szepannek, Thierry Monthé, and Claus Weihs	
Clustering Association Rules with Fuzzy Concepts	197
Matthias Steinbrecher and Rudolf Kruse	
Clustering with Repulsive Prototypes	207
Roland Winkler, Frank Rehm, and Rudolf Kruse	
Part III Mixture Analysis	
Weakly Homoscedastic Constraints for Mixtures of t-Distributions	219
Francesca Greselin and Salvatore Ingrassia	

Bayesian Methods for Graph Clustering	229
Pierre Latouche, Etienne Birmelé, and Christophe Ambroise	
Determining the Number of Components in Mixture Models for Hierarchical Data	241
Olga Lukočienė and Jeroen K. Vermunt	
Testing Mixed Distributions when the Mixing Distribution Is Known	251
Denys Pommeret	
Classification with a Mixture Model Having an Increasing Number of Components	261
Odile Pons	
Nonparametric Fine Tuning of Mixtures: Application to Non-Life Insurance Claims Distribution Estimation	271
Laure Sardet and Valentin Patilea	
Part IV Linguistics and Text Analysis	
Classification of Text Processing Components: The Tesla Role System	285
Jürgen Hermes and Stephan Schwiebert	
Nonparametric Distribution Analysis for Text Mining	295
Alexandros Karatzoglou, Ingo Feinerer, and Kurt Hornik	
Linear Coding of Non-linear Hierarchies: Revitalization of an Ancient Classification Method	307
Wiebke Petersen	
Automatic Dictionary Expansion Using Non-parallel Corpora	317
Reinhard Rapp and Michael Zock	
Multilingual Knowledge-Based Concept Recognition in Textual Data	327
Martin Schierle and Daniel Trabold	
Part V Pattern Recognition and Machine Learning	
A Diversified Investment Strategy Using Autonomous Agents	339
Rui Pedro Barbosa and Orlando Belo	
Classification with Kernel Mahalanobis Distance Classifiers	351
Bernard Haasdonk and Elżbieta Pełkalska	

Identifying Influential Cases in Kernel Fisher Discriminant Analysis by Using the Smallest Enclosing Hypersphere	363
Nelmarie Louw, Sarel Steel and Morne Lamont	
Self-Organising Maps for Image Segmentation	373
Ron Wehrens	
Image Based Mail Piece Identification Using Unsupervised Learning	385
Katja Worm and Beate Meffert	
Part VI Statistical Musicology	
Statistical Analysis of Human Body Movement and Group Interactions in Response to Music	399
Frank Desmet, Marc Leman, Micheline Lesaffre, and Leen De Bruyn	
Applying Statistical Models and Parametric Distance Measures for Music Similarity Search	409
Hanna Lukashevich, Christian Dittmar, and Christoph Bastuck	
Finding Music Fads by Clustering Online Radio Data with Emergent Self Organizing Maps	419
Florian Meyer and Alfred Ultsch	
Analysis of Polyphonic Musical Time Series	429
Katrin Sommer and Claus Weihs	
Part VII Banking and Finance	
Hedge Funds and Asset Allocation: Investor Confidence, Diversification Benefits, and a Change in Investment Style Composition	441
Wolfgang Bessler and Julian Holler	
Mixture Hidden Markov Models in Finance Research	451
José G. Dias, Jeroen K. Vermunt, and Sofia Ramos	
Multivariate Comparative Analysis of Stock Exchanges: The European Perspective	461
Julia Koralun-Bereźnicka	
Empirical Examination of Fundamental Indexation in the German Market	471
Max Mihm and Hermann Locarek-Junge	

The Analysis of Power for Some Chosen VaR Backtesting Procedures: Simulation Approach	481
Krzysztof Piontek	
Extreme Unconditional Dependence Vs. Multivariate GARCH Effect in the Analysis of Dependence Between High Losses on Polish and German Stock Indexes	491
Pawel Rokita	
Is Log Ratio a Good Value for Measuring Return in Stock Investments?	505
Alfred Ultsch	
Part VIII Marketing, Management Science and Economics	
Designing Products Using Quality Function Deployment and Conjoint Analysis: A Comparison in a Market for Elderly People	515
Samah Abu-Assab and Daniel Baier	
Analyzing the Stability of Price Response Functions: Measuring the Influence of Different Parameters in a Monte Carlo Comparison	527
Michael Brusch and Daniel Baier	
Real Options in the Assessment of New Products	537
Said Esber and Daniel Baier	
Exploring the Interaction Structure of Weblogs	545
Martin Klaus and Ralf Wagner	
Analyzing Preference Rankings when There Are Too Many Alternatives	553
Kar Yin Lam, Alex Koning, and Philip Hans Franses	
Considerations on the Impact of Ill-Conditioned Configurations in the CML Approach	563
A. Punzo	
Dyadic Interactions in Service Encounter: Bayesian SEM Approach	573
Adam Sagan and Magdalena Kowalska-Musiał	
Part IX Archaeology and Spatial Planning	
Estimating the Number of Buildings in Germany	585
M. Behnisch and A. Ultsch	

Mapping Findspots of Roman Military Brickstamps in <i>Mogontiacum</i> (Mainz) and Archaeometrical Analysis	595
Jens Dolata, Hans-Joachim Mucha, and Hans-Georg Bartel	
Analysis of Guarantor and Warrantee Relationships Among Government Officials in the Eighth Century in the Old Capital of Japan by Using Asymmetric Multidimensional Scaling	605
Akinori Okada and Towao Sakaehara	
Analysis of Massive Emigration from Poland: The Model-Based Clustering Approach	615
Ewa Witek	
Part X Bio- and Health Sciences	
Systematics of Short-Range Correlations in Eukaryotic Genomes	627
Jörn Hameister, Werner E. Helm, Marc-Thorsten Hütt, and Manuel Dehnert	
On Classification of Molecules and Species of Representation Rings	639
L. Häberle	
The Precise and Efficient Identification of Medical Order Forms Using Shape Trees	651
Uwe Henker, Uwe Petersohn, and Alfred Ultsch	
On the Prognostic Value of Gene Expression Signatures for Censored Data	663
Thomas Hielscher, Manuela Zucknick, Wiebke Werft, and Axel Benner	
Quality-Based Clustering of Functional Data: Applications to Time Course Microarray Data	675
Theresa Scharl and Friedrich Leisch	
A Comparison of Algorithms to Find Differentially Expressed Genes in Microarray Data	685
Alfred Ultsch, Christian Pallasch, Eckhard Bergmann, and Holger Christiansen	
Part XI Exploratory Data Analysis, Modeling and Applications	
Data Compression and Regression Based on Local Principal Curves	701
Jochen Einbeck, Ludger Evers, and Kirsty Hinchliff	

Optimization of Centrifugal Impeller Using Evolutionary Strategies and Artificial Neural Networks	713
René Meier and Franz Joos	
Efficient Media Exploitation Towards Collective Intelligence	723
Phivos Mylonas, Vassilios Solachidis, Andreas Geyer-Schulz, Bettina Hoser, Sam Chapman, Fabio Ciravegna, Steffen Staab and Pavel Smrz, Yiannis Kompatsiaris, and Yannis Avrithis	
Multi-class Extension of Verifiable Ensemble Models for Safety-Related Applications	733
Sebastian Nusser, Clemens Otte, and Werner Hauptmann	
Dynamic Disturbances in BTA Deep-Hole Drilling: Modelling Chatter and Spiralling as Regenerative Effects	745
Nils Raabe, Dirk Enk, Dirk Biermann, and Claus Weihs	
Nonnegative Matrix Factorization for Binary Data to Extract Elementary Failure Maps from Wafer Test Images	755
Reinhard Schachtner, Gerhard Pöppel, and Elmar Lang	
Collective Intelligence Generation from User Contributed Content	765
Vassilios Solachidis, Phivos Mylonas, Andreas Geyer-Schulz, Bettina Hoser, Sam Chapman, Fabio Ciravegna, Vita Lanfranchi, Ansgar Scherp, Steffen Staab, Costis Contopoulos, Ioanna Gkika, Byron Bakaimis, Pavel Smrz, Yiannis Kompatsiaris, and Yannis Avrithis	
Computation of the Molenaar Sijtsma Statistic	775
L. Andries van der Ark	
Keyword Index	785
Author Index	789