STUDIES IN CLASSIFICATION,
DATA ANALYSIS,
AND KNOWLEDGE ORGANIZATION

Data Analysis and Decision Support

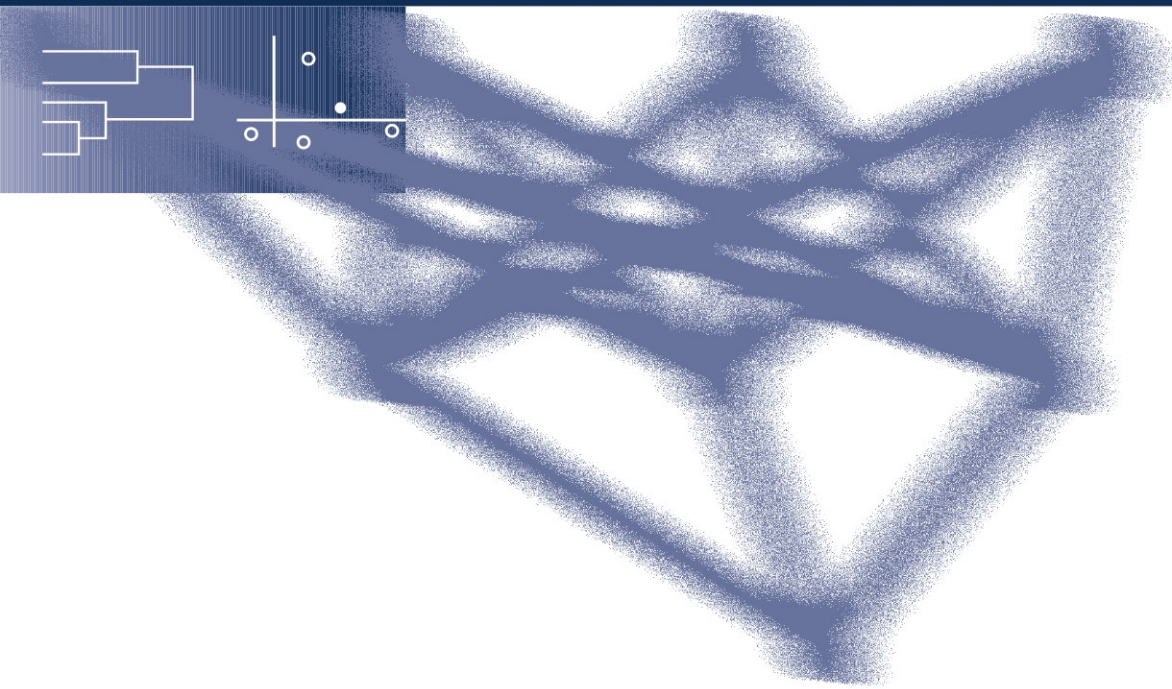
D. Baier
R. Decker
L. Schmidt-Thieme
Editors

Springer

STUDIES IN CLASSIFICATION,
DATA ANALYSIS,
AND KNOWLEDGE ORGANIZATION

D. Baier
R. Decker
L. Schmidt-Thieme
Editors

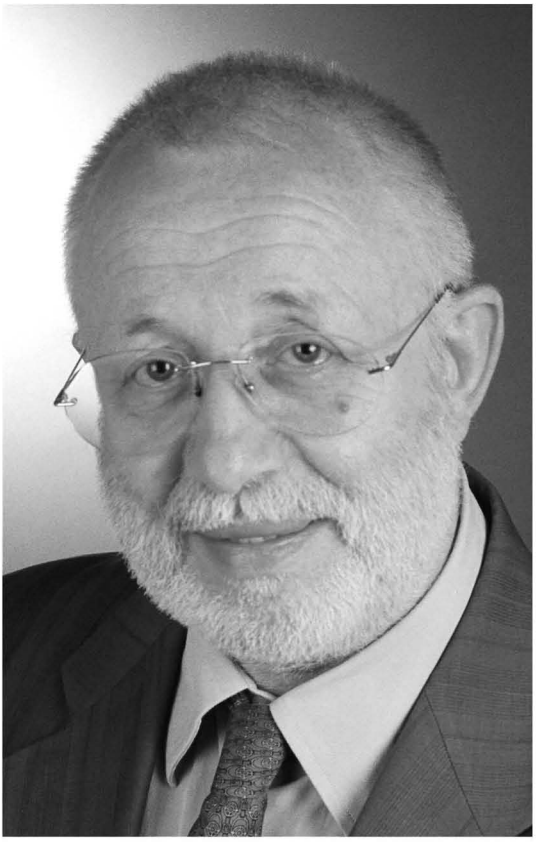
Data Analysis and Decision Support



Studies in Classification, Data Analysis, and Knowledge Organization

Managing Editors
H.-H. Bock, Aachen
W. Gaul, Karlsruhe
M. Vichi, Rome

Editorial Board
Ph. Arabie, Newark
D. Baier, Cottbus
F. Critchley, Milton Keynes
R. Decker, Bielefeld
E. Diday, Paris
M. Greenacre, Barcelona
C. Lauro, Naples
J. Meulman, Leiden
P. Monari, Bologna
S. Nishisato, Toronto
N. Ohsumi, Tokyo
O. Opitz, Augsburg
G. Ritter, Passau
M. Schader, Mannheim
C. Weihs, Dortmund



Wolfgang Gaul

Daniel Baier · Reinhold Decker
Lars Schmidt-Thieme
Editors

Data Analysis and Decision Support

Foreword by Shizuhiko Nishisato

Springer

Prof. Dr. Daniel Baier
Chair of Marketing and Innovation Management
Institute of Business Administration and Economics
Brandenburg University of Technology Cottbus
Konrad-Wachsmann-Allee 1
03046 Cottbus
Germany
daniel.baier@tu-cottbus.de

Prof. Dr. Reinhold Decker
Chair of Marketing
Department of Business Administration and Economics
Bielefeld University
Universitätsstr. 25
33615 Bielefeld
Germany
rdecker@wiwi.uni-bielefeld.de

Prof. Dr. Dr. Lars Schmidt-Thieme
Computer Based New Media Group (CGNM)
Institute for Computer Science
University of Freiburg
Georges-Köhler-Allee 51
79110 Freiburg
Germany
lst@informatik.uni-freiburg.de

ISSN 1431-8814
ISBN 3-540-26007-2 Springer-Verlag Berlin Heidelberg New York

Library of Congress Control Number: 2005926825

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer · Part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin · Heidelberg 2005
Printed in Germany

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Softcover-Design: Erich Kirchner, Heidelberg

SPIN 11427827 43/3153 – 5 4 3 2 1 0 – Printed on acid-free paper

Foreword

It is a great privilege and pleasure to write a foreword for a book honoring Wolfgang Gaul on the occasion of his sixtieth birthday. Wolfgang Gaul is currently Professor of Business Administration and Management Science and the Head of the Institute of Decision Theory and Management Science, Faculty of Economics, University of Karlsruhe (TH), Germany. He is, by any measure, one of the most distinguished and eminent scholars in the world today.

Wolfgang Gaul has been instrumental in numerous leading research initiatives and has achieved an unprecedented level of success in facilitating communication among researchers in diverse disciplines from around the world. A particularly remarkable and unique aspect of his work is that he has been a leading scholar in such diverse areas of research as graph theory and network models, reliability theory, stochastic optimization, operations research, probability theory, sampling theory, cluster analysis, scaling and multivariate data analysis. His activities have been directed not only at these and other theoretical topics, but also at applications of statistical and mathematical tools to a multitude of important problems in computer science (e.g., web-mining), business research (e.g., market segmentation), management science (e.g., decision support systems) and behavioral sciences (e.g., preference measurement and data mining). All of his endeavors have been accomplished at the highest level of professional excellence.

Wolfgang Gaul's distinguished contributions are reflected through more than 150 journal papers and three well-known books, as well as 17 edited books. This considerable number of edited books reflects his special ability to organize national and international conferences, and his skill and dedication in successfully providing research outputs with efficient vehicles of dissemination. His talents in this regard are second to none. His singular commitment is also reflected by his contributions as President of the German Classification Society, and as a member of boards of directors and trustees of numerous organizations and editorial boards. For these contributions, the scientific community owes him a profound debt of gratitude.

Wolfgang Gaul's impact on research has been felt in the lives of many researchers in many fields in many countries. The editors of this book, Daniel Baier, Reinhold Decker and Lars Schmidt-Thieme, are distinguished former students of Wolfgang Gaul, whom I had the pleasure of knowing when they were hard-working students under his caring supervision and guidance. This book is a fitting tribute to Wolfgang Gaul's outstanding research career, for

it is a collection of contributions by those who have been fortunate enough to know him personally and who admire him wholeheartedly as a person, teacher, mentor, and friend.

A glimpse of the content of the book shows two groups of papers, data analysis and decision support. The first section starts with symbolic data analysis, and then moves to such topics as cluster analysis, asymmetric multidimensional scaling, unfolding analysis, multidimensional data analysis, aggregation of ordinal judgments, neural nets, pattern analysis, Markov process, confidence intervals and ANOVA models with generalized inverses. The second section covers a wide range of papers related to decision support systems, including a long-term strategy for an urban transport system, loyalty programs, heuristic bundling, E-commerce, QFD and conjoint analysis, equity analysis, OR methods for risk management and German business cycles. This book showcases the tip of the iceberg of Wolfgang Gaul's influence and impact on a wide range of research. The editors' dedicated work in publishing this book is now amply rewarded.

Finally, a personal note. No matter what conferences one attends, Wolfgang Gaul always seems to be there, carrying a very heavy load of papers, transparencies, and a computer. He is always involved, always available, and always ready to share his knowledge and expertise. Fortunately, he is also highly organized - an important ingredient of his remarkable success and productivity. I am honoured indeed to be his colleague and friend.

Good teachers are those who can teach something important in life, and Wolfgang Gaul is certainly one of them. I hope that this book gives him some satisfaction, knowing that we all have learned a great deal from our association with him.

Toronto, Canada, April 2005 *Shizuhiko Nishisato*

Preface

This year, in July, Wolfgang Gaul will celebrate his 60th birthday. He is Professor of Business Administration and Management Science and one of the Heads of the Institute of Decision Theory and Management Science at the Faculty of Economics, University of Karlsruhe (TH), Germany. He received his Ph.D. and Habilitation in mathematics from the University of Bonn in 1974 and 1980 respectively.

For more than 35 years, he has been an active researcher at the interface between

- mathematics, operations research, and statistics,
- computer science, as well as
- management science and marketing

with an emphasis on data analysis and decision support related topics.

His publications and research interests include work in areas such as

- graph theory and network models, reliability theory, optimization, stochastic optimization, operations research, probability theory, statistics, sampling theory, and data analysis *(from a more theoretical point of view)* as well as
- applications of computer science, operations research, and management science, e.g., in marketing, market research and consumer behavior, product management, international marketing and management, innovation and entrepreneurship, pre-test and test market modelling, computer-assisted marketing and decision support, knowledge-based approaches for marketing, data and web mining, e-business, and recommender systems *(from a more application-oriented point of view).*

His work has been published in numerous journals like Annals of Operations Research, Applied Stochastic Models and Data Analysis, Behaviormetrika, Decision Support Systems, International Journal of Research in Marketing, Journal of Business Research, Journal of Classification, Journal of Econometrics, Journal of Information and Optimization Sciences, Journal of Marketing Research, Marketing ZFP, Methods of Operations Research, Zeitschrift für Betriebswirtschaft, Zeitschrift für betriebswirtschaftliche Forschung as well as in numerous refereed proceedings volumes.

His books on computer-assisted marketing and decision support – e.g. the well-known and wide-spread book "Computergestütztes Marketing" (published 1990 together with Martin Both) – imply early visions of the nowadays ubiquitous availability and usage of information-, model-, and knowledge-oriented decision aids for marketing managers. Equipped with a profound

mathematical background and a high degree of commitment to his research topics, Wolfgang Gaul has strongly contributed in transforming marketing and marketing research into a data-, model-, and decision-oriented quantitative discipline.

Wolfgang Gaul was one of the presidents of the German Classification Society GfKl (Gesellschaft für Klassifikation) and chaired the program committee of numerous international conferences. He is one of the managing editors of "Studies in Classification, Data Analysis, and Knowledge Organization", a series which aims at bringing together interdisciplinary research from different scientific areas in which the need for handling data problems and for providing decision support has been recognized. Furthermore, he was a scientific principal of comprehensive DFG projects on marketing and data analysis.

Last but not least Wolfgang Gaul has positively influenced the research interests and careers of many students. Three of them have decided to honor his merits with respect to data analysis and decision support by inviting colleagues and friends of him to provide a paper for this "Festschrift" and were delighted – but not surprised – about the positive reactions and the high number and quality of articles received.

The present volume is organized into two parts which try to reflect the research topics of Wolfgang Gaul: a more theoretical part on "Data Analysis" and a more application-oriented part on "Decision Support". Within these parts contributions are listed in alphabetical order with respect to the authors' names.

All authors send their congratulations

"Happy birthday, Wolfgang Gaul"

and hope that he will be as active in his and our research fields of interest in the future as he had been in the past.

Finally, the editors would like to cordially thank Dr. Alexandra Rese for her excellent work in preparing this volume, all authors for their cooperation during the editing process, as well as Dr. Martina Bihn and Christiane Beisel from Springer-Verlag for their help concerning all aspects of publication.

Cottbus, Bielefeld, Freiburg
April 2005

Daniel Baier
Reinhold Decker
Lars Schmidt-Thieme

Contents

Part I. Data Analysis

Optimization in Symbolic Data Analysis: Dissimilarities, Class Centers, and Clustering 3
Hans-Hermann Bock

An Efficient Branch and Bound Procedure for Restricted Principal Components Analysis 11
Wayne S. DeSarbo, Robert E. Hausman

A Tree Structured Classifier for Symbolic Class Description .. 21
Edwin Diday, M. Mehdi Limam, Suzanne Winsberg

A Diversity Measure for Tree-Based Classifier Ensembles 30
Eugeniusz Gatnar

Repeated Confidence Intervals in Self–Organizing Studies 39
Joachim Hartung, Guido Knapp

Fuzzy and Crisp Mahalanobis Fixed Point Clusters 47
Christian Hennig

Interpretation Aids for Multilayer Perceptron Neural Nets ... 57
Harald Hruschka

An Unfolding Scaling Model for Aggregated Preferential Choice Data 65
Tadashi Imaizumi

Model-Based Clustering – Discussion on Some Approaches ... 73
Krzysztof Jajuga

Three–Way Multidimensional Scaling: Formal Properties and Relationships Between Scaling Methods 82
Sabine Krolak–Schwerdt

Empirical Approach as a Scientific Framework for Data Analysis 91
Shizuhiko Nishisato

Asymmetric Multidimensional Scaling of Relationships Among Managers of a Firm 100
Akinori Okada, Tadashi Imaizumi, Hiroshi Inoue

Aggregation of Ordinal Judgements Based on Condorcet's Majority Rule 108
Otto Opitz, Henning Paul

ANOVA Models with Generalized Inverses 113
Wolfgang Polasek, Shuangzhe Liu

Patterns in Search Queries 122
Nadine Schmidt-Mänz, Martina Koch

Performance Drivers for Depth-First Frequent Pattern Mining 130
Lars Schmidt-Thieme, Martin Schader

On the Performance of Algorithms for Two-Mode Hierarchical Cluster Analysis – Results from a Monte Carlo Simulation Study 141
Manfred Schwaiger, Raimund Rix

Clustering Including Dimensionality Reduction 149
Maurizio Vichi

The Number of Clusters in Market Segmentation 157
Ralf Wagner, Sören W. Scholz, Reinhold Decker

On Variability of Optimal Policies in Markov Decision Processes 177
Karl-Heinz Waldmann

Part II. Decision Support

Linking Quality Function Deployment and Conjoint Analysis for New Product Design 189
Daniel Baier, Michael Brusch

Financial Management in an International Company: An OR-Based Approach for a Logistics Service Provider 199
Ingo Böckenholt, Herbert Geys

Development of a Long-Term Strategy for the Moscow Urban Transport System 204
Martin Both

The Importance of E-Commerce in China and Russia – An Empirical Comparison 212
Reinhold Decker, Antonia Hermelbracht, Frank Kroll

Analyzing Trading Behavior in Transaction Data of Electronic Election Markets 222
Markus Franke, Andreas Geyer-Schulz, Bettina Hoser

Critical Success Factors for Data Mining Projects 231
Andreas Hilbert

Equity Analysis by Functional Approach 241
Thomas Kämpke, Franz Josef Radermacher

A Multidimensional Approach to Country of Origin Effects in the Automobile Market 249
Michael Löffler, Ulrich Lutz

Loyalty Programs and Their Impact on Repeat Purchase Behaviour: An Extension on the "Single Source" Panel *BehaviorScan* 257
Lars Meyer-Waarden

An Empirical Examination of Daily Stock Return Distributions for U.S. Stocks 269
Svetlozar T. Rachev, Stoyan V. Stoyanov, Almira Biglova, Frank J. Fabozzi

Stages, Gates, and Conflicts in New Product Development: A Classification Approach 282
Alexandra Rese, Daniel Baier, Ralf Woll

Analytical Lead Management in the Automotive Industry 290
Frank Säuberlich, Kevin Smith, Mark Yuhn

Die Nutzung von multivariaten statistischen Verfahren in der Praxis - Ein Erfahrungsbericht 20 Jahre danach 300
Karla Schiller

Heuristic Bundling 313
Bernd Stauß, Volker Schlecht

The Option of No-Purchase in the Empirical Description of Brand Choice Behaviour 323
Udo Wagner, Heribert Reisinger

klaR Analyzing German Business Cycles 335
Claus Weihs, Uwe Ligges, Karsten Luebke, Nils Raabe

Index 345

Selected Publications of Wolfgang Gaul 347

Part I

Data Analysis

Optimization in Symbolic Data Analysis: Dissimilarities, Class Centers, and Clustering

Hans-Hermann Bock

Institut für Statistik und Wirtschaftsmathematik,
RWTH Aachen, Wüllnerstr. 3, D-52056 Aachen, Germany

Abstract. 'Symbolic Data Analysis' (SDA) provides tools for analyzing 'symbolic' data, i.e., data matrices $X = (x_{kj})$ where the entries x_{kj} are intervals, sets of categories, or frequency distributions instead of 'single values' (a real number, a category) as in the classical case. There exists a large number of empirical algorithms that generalize classical data analysis methods (PCA, clustering, factor analysis, etc.) to the 'symbolic' case. In this context, various optimization problems are formulated (optimum class centers, optimum clustering, optimum scaling,...). This paper presents some cases related to dissimilarities and class centers where explicit solutions are possible. We can integrate these results in the context of an appropriate k-means clustering algorithm. Moreover, and as a first step to probabilistically based results in SDA, we consider the definition and determination of set-valued class 'centers' in SDA and relate them to theorems on the 'approximation of distributions by sets'.

1 Symbolic data analysis

Classical data analysis considers single-valued variables such that, for n objects and p variables, each entry x_{kj} of the data matrix $X = (x_{kj})_{n\times p}$ is a real number (quantitative case) or a category (qualitative case). The term *symbolic data* relates to more general scenarios where x_{kj} may be an interval $x_{kj} = [a_{kj}, b_{kj}] \in \mathbb{R}$ (e.g., the interquartile interval of fuel prices in a city), a set $x_{kj} = \{\alpha, \beta, ...\}$ of categories (e.g., $\{green, red, black\}$ the favourite car colours in 2003), or even a frequency distribution (the histogram of monthly salaries in Karlsruhe in 2000). Various statistical methods and a software system SODAS have been developed for the analysis of symbolic data (see Bock and Diday (2000)). In the context of these methods, there arise various mathematical optimization problems, e.g., when defining the dissimilarity between objects (intervals in $\mathbb{R}^p$), when characterizing a 'most typical' cluster representative (class center), and when defining optimum clusterings.

This paper describes some of these optimization problems where a more or less explicit solution can be given. We concentrate on the case of *interval-type data* where each object $k = 1, ..., n$ is characterized by a data vector $x_k = ([a_{k1}, b_{k1}], ..., [a_{kp}, b_{kp}])$ with component-specific intervals $[a_{kj}, b_{kj}] \in \mathbb{R}$. Such data can be viewed as n p-dimensional intervals (rectangles, hypercubes) $Q_1, ..., Q_n \subset \mathbb{R}^p$ with $Q_k := [a_{k1}, b_{k1}] \times \cdots \times [a_{kp}, b_{kp}]$ for $k = 1, ..., n$.