

C. Hayashi · N. Ohsumi
K. Yajima · Y. Tanaka
H.-H. Bock · Y. Baba (Eds.)

Data Science, Classification, and Related Methods

Proceedings of the Fifth Conference of the International
Federation of Classification Societies (IFCS-96),
Kobe, Japan, March 27–30, 1996

With 240 Figures



Springer

Prof. Emeritus Chikio Hayashi
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Prof. Keiji Yajima
School of Management, Science University of Tokyo
500 Shimokiyoku, Kuki, Saitama 346, Japan

Prof. Hans-Hermann Bock
Institut für Statistik
Rheinisch-Westfälische Technische Hochschule (RWTH)
D-52056 Aachen, Germany

Prof. Noboru Ohsumi
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Prof. Yutaka Tanaka
Faculty of Environmental Science & Technology, Okayama University
2-1-1 Tsushima-naka, Okayama 700, Japan

Prof. Yasumasa Baba
The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

ISBN 978-4-431-70208-5 ISBN 978-4-431-65950-1 (eBook)
DOI 10.1007/978-4-431-65950-1

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 microfilms or in any other ways, and storage in data banks.

© Springer Japan 1998

Originally published by Springer-Verlag Tokyo Berlin Heidelberg New York in 1998

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.

Product liability: The publishers cannot guarantee the accuracy of any information about the application of operative techniques and medications contained in this book. In every individual case the user must check such information by consulting the relevant literature.

TABLE OF CONTENTS

Preface	V
Conference Committee	VIII
Local Editorial Board	IX
List of Reviewers	IX

Part I: General Aspects of Data Science

Probabilistic Aspects in Classification	
<i>Hans H. Bock</i>	3
Cluster Validation	
<i>A. D. Gordon</i>	22
What is Data Science? – Fundamental Concepts and a Heuristic Example	
<i>Chikio Hayashi</i>	40
Fitting Graphs and Trees With Multidimensional Scaling Methods	
<i>Willem J. Heiser</i>	52
Classification and Data Analysis in Finance	
<i>Krzysztof Jajuga</i>	63
How to Validate Phylogenetic Trees?: A Stepwise Procedure	
<i>François-Joseph Lapointe</i>	71
Some Trends in the Classification of Variables	
<i>F. Costa Nicolau and H. Bacelar-Nicolau</i>	89
Convexity Methods in Classification	
<i>Jean-Paul Rasson</i>	99

Part II: Methodologies in Classification

Evaluation and Assessment Procedures

How Many Clusters? – An Investigation of Five Procedures for Detecting Nested Cluster Structure	
<i>A. D. Gordon</i>	109
Partitional Cluster Analysis With Genetic Algorithms: Searching for the Number of Clusters	
<i>J.A. Lozano, P. Larrañaga, and M. Graña</i>	117
Explanatory Variables in Classifications and the Detection of the Optimum Number of Clusters	
<i>János Podani</i>	125

Random Dendrograms for Classifiability Testing <i>Bernard Van Cutsem, Bernard Ycart</i>	133
The L_p -Product of Ultrametric Spaces and the Corresponding Product of Hierarchies <i>Bernard Fichet</i>	145
Towards Comparison of Decomposable Systems <i>Mark Sh. Levin</i>	154
Performance of Eight Dissimilarity Coefficients to Cluster a Compositional Data Set <i>Maria Cristina Martín</i>	162
Topics in Clustering and Classification	
Consensus of Hierarchical Classifications <i>Bruno Simeone and Maurizio Vichi</i>	170
On the Minimum Description Length (MDL) Principle for Hierarchical Classifications <i>Peter G. Bryant</i>	182
Consensus Methods for Pyramids and Other Hypergraphs <i>J. Lehel, F. R. McMorris, and R. C. Powers</i>	187
On the Behavior of Splitting Criteria for Classification Trees <i>Roberta Siciliano and Francesco Mola</i>	191
Fitting Pre-Specified Blockmodels <i>Vladimir Batagelj, Anuška Ferligoj, and Patrik Doreian</i>	199
Robust Impurity Measures in Decision Trees <i>Tomàs Aluja-Banet and Eduard Nafria</i>	207
Induction of Decision Trees Based on the Rough Set Theory <i>Tu Bao Ho, Trong Dung Nguyen, and Masayuki Kimura</i>	215
Visualizing Data in Tree—Structured Classification <i>Francesco Mola and Roberta Siciliano</i>	223
Adaptive Cluster Analysis Techniques – Software and Applications <i>Hans-Joachim Mucha, Rainer Siegmund-Schultze, and Karl Dübön</i>	231

Part III: Classification and Discrimination

Statistical Approaches for Classification Problems

The Most Random Partition of a Finite Set and Its Application to Classification <i>Masaaki Sibuya</i>	241
A Mixture Model to Classify Individual Profiles of Repeated Measurements <i>Toshiro Tango</i>	247
Irregularly Spaced AR(ISAR) Models <i>Jeffrey S.C. Pai, Wolfgang Polasek, and Hideo Kozumi</i>	255

Discrimination and Pattern Analysis

Two Types of Partial Least Squares Method in Linear Discriminant Analysis <i>Hyun Bin Kim and Yutaka Tanaka</i>	261
Resampling Methods for Error Rate Estimation in Discriminant Analysis <i>Masayuki Honda and Sadanori Konishi</i>	268
A Short Overview of the Methods for Spatial Data Analysis <i>Masaharu Tanemura</i>	276

Choice of Multiple Representative Signatures for On-Line Signature Verification Using a Clustering Procedure <i>Isao Yoshimura, Mitsu Yoshimura, and Shin-ichi Matsuda</i>	284
--	-----

Part IV: Related Approaches for Classification

Fuzzy and Probabilistic Modeling Methods

Algorithms for L_1 and L_p Fuzzy c-Means and Their Convergence <i>Sadaaki Miyamoto and Yudi Agusta</i>	295
General Approach to the Construction of Measures of Fuzziness of Fuzzy K-Partitions <i>Slavka Bodjanova</i>	303
Additive Clustering Model and Its Generalization <i>Mika Sato and Yoshiharu Sato</i>	312
A Proposal of an Extended Model of ADCLUS Model <i>Tadashi Imaizumi</i>	320

Spatial Clustering and Neural Networks

Comparison of Pruning Algorithms in Neural Networks <i>Yoshihiko Hamamoto, Toshinori Hase, Satoshi Nakai, and Shingo Tomita</i>	328
Classification Method by Using the Associative Memories in Cellular Neural Networks <i>Akihiro Kanagawa, Hiroaki Kawabata, and Hiromitsu Takahashi</i>	334
Application of Kohonen Maps to the GPS Stochastic Tomography of the Ionosphere <i>M. Hernández-Pajares, J. M. Juan, and J. Sanz</i>	341

Symbolic and Conceptual Data Analysis

Capacities, Credibilities in Analysis of Probabilistic Objects by Histograms and Lattices <i>Edwin Diday and Richard Emilion</i>	353
Symbolic Pattern Classifiers Based on the Cartesian System Model <i>Manabu Ichino and Hiroyuki Yaguchi</i>	358
Extension Based Proximities Between Constrained Boolean Symbolic Objects <i>Francisco de A. T. de Carvalho</i>	370
Towards a Normal Symbolic Form <i>Marc Csernel and Francisco de A.T. de Carvalho</i>	379
The SODAS Project : A Software for Symbolic Data Analysis <i>Georges Hébrail</i>	387
Classification Structures for Cognitive Maps <i>Stephen C. Hirtle and Guoray Cai</i>	394
Unsupervised Concept Learning Using Rough Concept Analysis <i>Tu Bao Ho</i>	404
Implicative Statistical Analysis <i>R. Gras, H. Briand, P. Peter, and J. Philippe</i>	412

Part V: Correspondence Analysis, Quantification Methods, and Multidimensional Scaling

Correspondence Analysis and Its Application

Correspondence Analysis, Discrimination, and Neural Networks <i>Ludovic Lebart</i>	423
---	-----

Exploratory Data Analysis for Hayashi's Quantification Method III by Graphics <i>Tsutomu Komazawa and Takahiro Tsuchiya</i>	431
Exploring Multidimensional Quantification Space <i>Shizuhiko Nishisato</i>	441
Homogeneity Analysis for Partitioning Qualitative Variables <i>Takahiro Tsuchiya</i>	452
Determining the Distance Index, II <i>Matevž Bren and Vladimir Batagelj</i>	460
Classification of Textual Data	
Meta-Data and Strategies of Textual Data Analysis: Problems and Instruments <i>Sergio Bolasco</i>	468
Clustering of Texts Using Semantic Graphs. Application to Open-Ended Questions in Surveys <i>Monica Bécue Bertaut and Ludovic Lebart</i>	480
How to Find the Nearest by Evaluating Only Few? —Clustering Techniques Used to Improve the Efficiency of an Information Retrieval System Based on Distributional Semantics <i>Martin Rajman and Arnon Rungsawang</i>	488
Multidimensional Scaling	
Fitting the CANDCLUS / MUMCLUS Models with Partitioning and Other Constraints <i>J. Douglas Carroll and Anil Chaturvedi</i>	496
A Distance-Based Biplot for Multidimensional Scaling of Multivariate Data <i>Jacqueline J. Meulman</i>	506
Latent-Class Scaling Models for the Analysis of Longitudinal Choice Data <i>Ulf Böckenholt</i>	518

Part VI: Multivariate and Multidimensional Data Analysis

Multidimensional Data Analysis

Nonlinear Multivariate Analysis by Neural Network Models <i>Yoshio Takane</i>	527
Generalized Canonical Correlation Analysis with Linear Constraints <i>Haruo Yanai</i>	539
Principal Component Analysis Based on a Subset of Variables for Qualitative Data <i>Yuichi Mori, Yutaka Tanaka, and Tomoyuki Tarumi</i>	547
Missing Data Imputation in Multivariate Analysis <i>Mario Romanazzi</i>	555

Multiway Data Analysis

Recent Developments in Three-Mode Factor Analysis — Constrained Three-Mode Factor Analysis and Core Rotations <i>Henk A. L. Kiers</i>	563
Tucker2 as a Second-Order Principal Component Analysis <i>Takashi Murakami</i>	575
Parallel Factor Analysis with Constraints on the Configurations : An Overview <i>Pieter M. Kroonenberg and Willem J. Heiser</i>	587

Non-Linear Modeling and Visual Treatment

Regression Splines for Multivariate Additive Modeling <i>Jean-Francois Durand</i>	598
Bounded Algebraic Curve Fitting for Multidimensional Data Using the Least-Squares Distance <i>Masahiro Mizuta</i>	610
Using the Wavelet Transform for Multivariate Data Analysis and Time Series Analysis <i>Fionn Murtagh and Alexandre Aussem</i>	617
Visual Manipulation Environment for Data Analysis System <i>Masahiro Mizuta and Hiroyuki Minami</i>	625
Human Interface for Multimedia Database With Visual Interaction Facilities <i>Toshikazu Kato</i>	632

Part VII: Case Studies of Data Science

Social Science and Behavioral Science

Proposition of a New Paradigm for a Scientific Classification for Leadership Behavior Research Perspective <i>Jyuji Misumi</i>	647
Structure of Attitude Toward Nuclear Power Generation <i>Kiyoshi Karube, Chikio Hayashi, and Shinichi Morikawa</i>	653
Research Concerning the Consciousness of Women's Attitude Toward Independence <i>Setsuko Takakura</i>	661
A Cross-National Analysis of the Relationship Between Genderedness in the Legal Naming of Same-Sex Sexual/Intimate Relationships and the Gender System <i>Saori Kamano</i>	669

Management Science and Marketing Science

A Constrained Clusterwise Regression Procedure for Benefit Segmentation <i>Daniel Baier</i>	676
Application of Classification and Related Methods to SQC Renaissance in Toyota Motor <i>Kakuro Amasaka</i>	684
Application of Statistical Binary Tree Analysis to Quality Control <i>Atsushi Ootaki</i>	696
Analysis of Preferences for Telecommunication Services in Each Area <i>Tohru Ueda and Daisuke Satoh</i>	708
Effects of End-Aisle Display and Flier on the Brand-Switching of Instant Coffee <i>Akinori Okada</i>	716

Environmental, Ecological, Biological, and Medical Sciences

On the Classification of Environmental Data in the Bavarian Environment Information System Using an Object-Oriented Approach <i>Erich Weihs</i>	728
Cluster Analysis of Associated Words Obtained From a Free Response Test on Tokyo Bay <i>Shinsuke Suga, Ko Oi, and Sadaaki Miyamoto</i>	736

Data Analysis of Deer-Train Collisions in Eastern Hokkaido, Japan <i>Keiichi Onoyama, Noboru Ohsumi, Naoko Mitsumochi, and Tsuyoshi Kishihara</i>	746
Comparison of Some Numerical Data Between the <i>belisama</i> Group of the Genus <i>Delias</i> Hübner (Insecta: Lepidoptera) From Bali Island, Indonesia <i>Sadaharu Morinaka</i>	752
A Method for Classifying Unaligned Biological Sequences <i>B. Tallur and J. Nicolas</i>	758
An Approach to Determine the Necessity of Orthognathic Osteotomy or Orthodontic Treatment in a Cleft Individual – Comparison of Craniomaxillo-Facial Structures in Borderline Cases by Roentgenocephalometrics <i>Sumimasa Ohtsuka, Fumiye Ohmori, Kazunobu Imamura, Yoshinobu Shibasaki, and Noboru Ohsumi</i>	766
Data Analysis for Quality of Life and Personality <i>Kazue Yamaoka and Mariko Watanabe</i>	774
Contributor Index	779