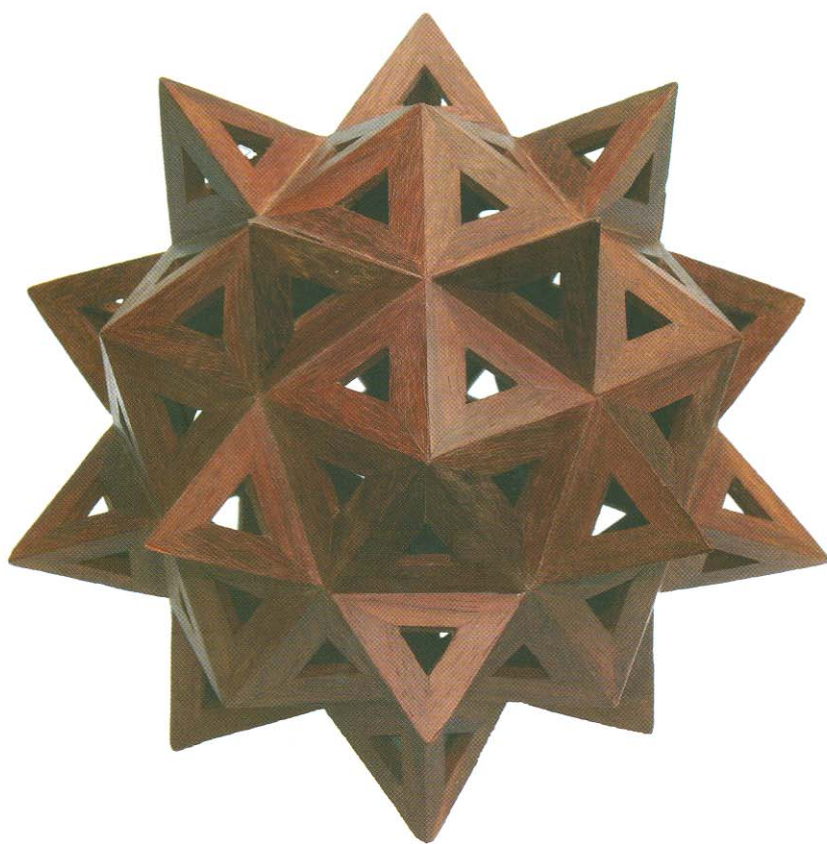




Classification and Data Analysis 2007

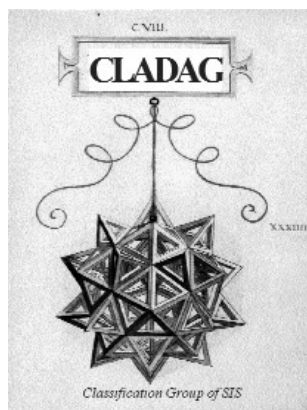
Book of Short Papers

Meeting of the Classification and Data Analysis
Group of the Italian Statistical Society

eum > economia > statistica

CLASSIFICATION AND DATA ANALYSIS 2007

Book of Short Papers
Meeting of the
CLAssification and Data Analysis Group
of the Italian Statistical Society



MACERATA - SEPTEMBER, 12TH-14TH, 2007

eum

Isbn 978-88-6056-020-9

Prima edizione: settembre 2007

© 2007 eum edizioni università di macerata

Vicolo Tornabuoni, 58 - 62100 Macerata

info.ceum@unimc.it

<http://ceum.unimc.it>

Realizzazione e distribuzione:

Quodlibet società cooperativa

Via S. Maria della Porta, 43 - 62100 Macerata

www.quodlibet.it

Stampa: Grafica Edirice Romana s.r.l., Roma

This book was realized with the financial support of



Università degli Studi
di Macerata



REGIONE
MARCHE



SISTEMA
INFORMATIVO
STATISTICO

Table of contents

<i>Preface</i>	19
----------------	----

KEY-NOTE PAPERS

S. R. Masera, G. Mazzoni <i>Actuarial and Continuous Time Risk Models: Towards a Synthesis, Changing financial paradigms in the new enterprise economy</i>	23
G. McLachlan <i>Clustering of High-Dimensional and Correlated Data</i>	29
A. Montanari <i>Classification by mixture and latent variable models</i>	37
A. Rizzi <i>Statistical Methods for Cryptography</i>	43
R. Zang, H. Bozdogan <i>Model Selection in Relevance Vector Machines (RVMs) Using Information Complexity and Genetic Algorithm (GA)</i>	51

INVITED PAPERS

Specialized Session 1

Knowledge extraction from temporal data models

Organizer: D. Piccolo

R. Baragona, S. Vitrano <i>Statistical and numerical algorithms for time series classification</i>	65
M. Corduas <i>Comparing time series: shape-based or structural similarities?</i>	69
G. Scepi, G. Milone <i>Temporal Data Mining: clustering methods and algorithms</i>	73

Specialized Session 2

Statistical models with errors-in-covariates

Organizer: A. Pastore

M. Battauz, R. Bellio	
<i>Structural analysis of linear mixed models with measurement error</i>	79
A. Guolo	
<i>Measurement error correction techniques in case-control studies</i>	83
P. Mantovan, A. Pastore	
<i>Dynamic regression with covariate errors by covariate local clusters</i>	87

Specialized Session 3

Multivariate Analysis for Microarray Data

Organizer: M Vichi

M. Alfò, F. Martella, M. Vichi	93
<i>Biclustering of microarray data</i>	
A. M. Mineo, L. Augugliaro, C. Fede, M. Ruggieri	
<i>A Statistical Calibration Method Based on Non-Linear Mixed Model for Affymetrix Probe Level Data</i>	97
E. Wit, G. Green	
<i>Random effects modelling for multivariate data from cDNA microarrays</i>	101

Specialized Session 4

Cluster Analysis of complex data

Organizer: R. Verde

A. Irpino, R. Verde	
<i>Clustering linear models using Wasserstein distance</i>	107
F. Palumbo, A. Iodice D'Enza	
<i>A two-step iterative procedure for clustering of binary sequences</i>	111
A. Petrucci, C. T. Brownless	
<i>Spatial Clustering Methods for the Detection of Homogenous Areas</i>	115

Specialized Session 5

Educational Processes Assesment by means of Latent Variables Models

Organizer: S. Mignani

M. Battauz , E. Gori	
<i>Educational assessment in presence of heteroscedastic measurement errors</i>	121

B. Chiandotto, F. Polverini, B. Bertaccini
The effectiveness of university education: a structural equation models 125

W. J. van Der Linden
Response Times on Test Items: Models and Applications 129

Specialized Session 6
Classification of complex data
 Organizer: D.A. Zighed

H. Azzag, C. Guinot, G. Venturini
On data clustering with bio-inspired algorithms 135

F. de Carvalho, Y. Lechevallier, R. Verde
Clustering approach on interval data 139

B. Fichet
Around the ultrametric sandwich 143

D. A. Zighed
Separability of classes in a multidimensional space 147

Specialized Session 7
Multidimensional Scaling
 Organizer: A. Okada

G. Bove
Models for asymmetry in proximity data 153

T. Imaizumi
On extracting a local structure of asymmetric matrix 157

A. Okada
Two-Dimensional Centrality of Asymmetric Social Network 161

Specialized Session 8
Statistical Models for Public Policies
 Organizer: S. Ingrassia

M. Caserta
Enlarging internal regional markets: measuring the effects on local development 167

R. Innocenti, A. Giommi, C. Brownless
A New Statistical Zoning for the Municipality of Firenze 171

M. Riani, R. Lagomarsini, A. Micozzi
Robust Clustering for Performance Evaluation 175

G. Vittadini, S.C. Minotti, M. Sanarico	
<i>Cluster-Weighted Modeling for Evaluating the Relative Effectiveness of Healthcare Institutions</i>	179

Specialized Session 9

Classification models for enterprise risk management

Organizer: P. Giudici

E. Bonafede, P. Cerchiello	
<i>A proposal to fuzzify categorical variables in operational risk management</i>	185
C. Cornalba	
<i>Statistical models to measure IT operational risks</i>	189
D. Fantazzini, S. Figini	
<i>Predictive dynamic models for SMEs</i>	193

Specialized Session 10

Model based clustering

Organizer: D. Vicari

C. Biernacki, A. Lourme	
<i>Simultaneous Model-Based Clustering of Data Arising from Different Populations</i>	199
G. Galimberti, G. Soffritti	
<i>Multiple cluster structures and mixture models: recent developments for multilevel data</i>	203
D. Vicari, M. Vichi	
<i>Multivariate regression model with clustering of objects and variables</i>	207

CONTRIBUTED PAPERS

Solicited Papers Session 1

Applications in Classification and Segmentation

Organizer: J. Antoch

J. Antoch	
<i>Testing the Difference of the ROC Curves in Bigamma Model</i>	217
C. Conversano, E. Dusseldorp	
<i>Classification Trunk Approach for Variable Selection and Threshold Interactions</i>	219

J. Klaschka	
<i>Tree-based classification of EEG spectra</i>	223
R. Miele	
<i>An Ant Colony-based Algorithm for Classification and Regression Trees"</i>	227
R. Siciliano, V. A. Tutore, M. Aria	
<i>3Way Trees</i>	231

Solicited Papers Session 2

Classification Issues in Social Network Analysis

Organizer: G. Giordano, M. P. Vitale

V. Batagelj	
<i>Clustering in Networks</i>	237
G. Giordano, M. P. Vitale	
<i>Local Factorial Analysis and Contiguity Constraints in Social Networks</i>	241
Y. H. Said, E. J. Wegman, W. Sharabati, J. T. Rigsby	
<i>Social Networks of Author-Coauthor Relationships</i>	245
S. Zaccarin, G. Rivellini	
<i>Modelling network data: an Introduction to Exponential Random Graph Models</i>	249

Solicited Papers Session 3

Metainformation and knowledge extraction from textual data bases

Organizer: S. Balbi

S. Bolasco, P. Pavone	
<i>Automatic dictionary and rule-based systems for extracting information from text</i>	255
A. Canzonetti	
<i>Semantic classification and cooccurrences: a method for the rules production for the information extraction from textual data</i>	259
F. della Ratta, B. Lorè, G. La Rocca	
<i>Textual analysis perspectives on categorisation of activities in Istat survey on occupations</i>	263
G. Infante, M. Misuraca	
<i>Text Mining Strategies for Analyzing Semi-structured Corpora</i>	267

Solicited Papers Session 4

Customer Relationship Management

Organizer: C. Davino

- F. Camillo, C. Liberati
A micro-data mining approach for qualitative-emotional marketing using neuro-information 273
- C. Davino, R. Del Gobbo
Structural Neural Networks for Modeling Customer Satisfaction 277
- S. Figini, A. Roccato
How to improve predictive models for database marketing applications 281

Solicited Papers Session 5

Missing Data

Organizer: A. Plaia

- M. Aria, A. D'Ambrosio, R. Siciliano
Robust Incremental Trees for Missing Data Imputation and Data Fusion 287
- A. Plaia, A. L. Bondi
Regression imputation for Space-Time datasets with missing values 291
- I. Sulis, M. Porcu
A multiple imputation approach in a survey on university teaching evaluation 295

Solicited Papers Session 6

Statistical methods in decision-making

Organizer: R. Amenta

- G. Adelfio, M. Chiodi, D. Luzio
An algorithm for earthquakes clustering based on maximum likelihood 301
- F. Di Salvo, N. Ferotti, S. Consagra
Hospital performance comparison: assessing inappropriate stay in the hospitals of Palermo 305
- C. Gaetan, L. Greco
Weighted Likelihood Inference in Gaussian Spatial Linear Models 309

Solicited Papers Session 7

Multivariate methods in social research

Organizer: D. F. Iezzi

- M. Civardi, F. Crippa
From University to Work: a measure of coherence between education and job 315

P. Costantini	
<i>Analyzing learning effects through Latent Growth Models</i>	319
L. Fabbri	
<i>Dimensionality of scores obtained with a pair-comparison tournament system of questionnaire items</i>	323
D. F. Iezzi, M. Grisoli	
<i>Using Rasch measurement to assess the role of the traditional family in Italy</i>	327

Solicited Papers Session 8

Statistical Models with Latent Features

Organizer: L. Tardella

S. Arima, L. Tardella	
<i>A Bayesian Approach to Peabody Picture Vocabulary Test Revised</i>	333
F. Bartolucci, A. Farcomeni	
<i>Dynamic logit models for panel data based on a latent Markov heterogeneity structure</i>	337
G. Petris, S. Petrone	
<i>Bayesian inference for dynamic linear models with random variances</i>	341
R. Rocci, A. Maruotti	
<i>An INDSCAL based mixture model</i>	345

Solicited Papers Session 9

Classification Trees

Organizer: R. Siciliano, M. Aria

C. Conversano, F. Mola	
<i>Sequential Automatic Search of a Subset of Classifiers for Multi-Attribute Response Prediction</i>	351
G. Galimberti, M. Pillati, G. Soffritti	
<i>Comparing strategies for robust regression tree construction</i>	355
V. A. Tutore, A. Mooijart	
<i>Optimal Scaling Trees</i>	359
M. Vezzoli, C. Stone	
<i>CRAGGING</i>	363

Solicited Papers Session 10

Exploring data structures with the Forward Search

Organizer: A. Cerioli

M. Bini, B. Bertaccini	
<i>Robust fuzzy classification</i>	369
D. G. Calò	
<i>Outlier detection via Forward Search: a proposal based on mixture models</i>	373
D. Perrotta, F. Torti	
<i>Detecting price outliers in European trade data with backward and forward methods</i>	377
N. Solaro, M. Pagani	
<i>The Forward Search for Metric Multidimensional Scaling</i>	381

Contributed Papers Session 1

Clustering in Marketing Research and Data Mining

R. Menichelli, F. Palumbo	
<i>Clustering of binary data in AR mining: a comparison among different strategies</i>	387
I. Morlini, S. Zani	
<i>Comparing approaches for clustering mixed mode data: an application in marketing research</i>	391
G. Schoier, B. Bato	
<i>A modification of the DBSCAN Algorithm in a Spatial Data Mining Approach</i>	395
A. Tarsitano, I. Bonafine	
<i>Distance function for mixed type data</i>	399
A. F. X. Wilhelm	
<i>Association rule mining of multimedia content</i>	403

Contributed Papers Session 2

Data Analysis in Finance

A. Amendola, M. Niglio, C. Vitale	
<i>Temporal aggregation and closure of VARMA models. Some new results</i>	409
L. Attardi, D. Vistocco	
<i>Evaluating financial portfolio activeness through style analysis models</i>	413

R. Castellano, L. Scaccia	
<i>Bayesian hidden Markov models for financial data</i>	417
S. Cieply, R. Abdesselam	
<i>Interrelations between Asset and Financial Structures: the case of French NTIC firms</i>	421
F. Domma, S. Giordano, P. F. Perri	
<i>Modelling the dependence structure of financial assets using copula functions</i>	425
M. Marozzi, L. Santamaria	
<i>How to Make a Complex Corporate Finance Indicator Simpler</i>	429

Contributed Papers Session 3 *Mixture and Latent Class model*

M. Alfò, A. Maruotti	
<i>A Hierarchical Model for Time Dependent Longitudinal Data</i>	435
P. Coretto, C. Hennig	
<i>Choice of the improper density in robust improper ML for finite normal mixture</i>	439
G. E. Montanari, M. G. Ranalli, P. Eusebi	
<i>Multilevel Latent Class Models for evaluation of long term care facilities</i>	443
R. Piccarreta, M. Bonetti, S. Venturini	
<i>Clustering-based measurement of dependence</i>	447
L. Scrucca	
<i>Visualization of model-based clustering structures</i>	451
K. Yamaguchi, M. Watanabe	
<i>Mixture models for multivariate count data</i>	455

Contributed Papers Session 4 *Economics*

V. Adorno, C. Bernini, G. Pellegrini	
<i>Comparing matching methods in policy evaluation</i>	461
C. Cappelli, F. Di Iorio	
<i>Structural breaks versus long memory, a simulation study</i>	465
D. De Vos, K. Vanhoof, H. Van Landeghem	
<i>Symbolic Data As Measures For Logistical Performances</i>	469
C. Giusti	
<i>Multiple imputation of income for the Labour Force Survey</i>	473

Contributed Papers Session 5

Latent Variables

G. Adelfio	
<i>Earthquakes clustering interpretation based on a second-order diagnostic approach</i>	479
S. Bianconcini, S. Cagnone	
<i>A Multilevel Latent Variable Model for Multidimensional Longitudinal Data</i>	483
E. Otranto	
<i>A time varying Hidden Markov Model with latent information</i>	487

Contributed Papers Session 6

Non Parametric Regression

D. Coin	
<i>A test for Normality with high Power against Symmetrical Alternatives</i>	493
L. De Benedictis, M. Callegati, M. Tamberi	
<i>Semiparametric analysis of the specialization-income relationship</i>	497
R. Miglio	
<i>Non Parametric Procedures to Explore Interactions in Epidemiological Studies</i>	501
A. Zirilli, A. Alibrandi	
<i>SMART Regression: an application to multidimensional data in a medical context</i>	505

Contributed Papers Session 7

Cluster Structures

A. Alibrandi, M. Giacalone	
<i>A Comparison among Statistical Criteria to Establish the Number of Clusters in Multidimensional Data</i>	511
A. Balzanella, E. Romano, R. Verde	
<i>Outliers detection for a Robust Curves Clustering Algorithm</i>	515
A. Iodice D'Enza	
<i>Visual Monitoring Tool of Association Patterns in Binary Data Flows</i>	519
G. Notarstefano, R. Scuderi	
<i>Polarization and Dynamic Clustering: A Review</i>	523
S. Polettini, L. Di Consiglio	
<i>Disclosure risk estimation with survey data: a comparative study</i>	527

Contributed Papers Session 8

Customer Satisfaction

L. Berzieri	
<i>Demographic snapshot of Parma electoral colleges</i>	533
P. Chirico	
<i>Common Optimal Scaling for Customer Satisfaction Multidimensional Models</i>	537
M. Civardi, E. Zavarrone	
<i>Ordinal Scales and Customer Satisfaction: an Applet</i>	541
P. Ferrari, C. V. Fiorio, L. Pagani	
<i>A two-step procedure to analyze users' satisfaction</i>	545

Contributed Papers Session 9

Discrimination

R. Abdesselam	
<i>Mixed Discriminant Analysis</i>	551
E. Bricchieri Colombi, C. T. Brownlees, G. M. Gallo	
<i>Predicting Binary Outcomes Using Flexible and Parsee's Methods</i>	555
S. Ingrassia, I. Morlini	
<i>Computational experiences with the equivalent number of degrees of freedom of neural networks</i>	559
R. Sanzo, E. Trinca	
<i>A multivariate approach to non-response treatment for Foreign control companies Survey</i>	563

Contributed Papers Session 10

Robust Classification

A. Corbellini, P. Ganugi, L. Crosato	
<i>Robust sequential Fitting of the Pareto II distribution: Theoretical and Computational Aspects</i>	569
M. Daniele, A. Bondi	
<i>Outliers in hierarchical models: an application to air pollution data</i>	573
A. Gottard, S. Pacillio	
<i>Robustifying SINful procedure</i>	577
P. Mascia, R. Miele, F. Mola	
<i>Robustness Issues in Classification and Regression Trees</i>	581
S. Scippacercola	
<i>The Progressive Single Linkage Algorithm Based on Minkowski Ultrametrics</i>	585

Contributed Papers Session 11

Bayesian Network

A. Brogini, D. Slanzi	
<i>Several computational studies about variable selection for Bayesian networks</i>	591
M. Mascherini	
<i>Exploring eBusiness Readiness Indicators using Bayesian Networks</i>	595
L. Scaccia, F. Bartolucci	
<i>Bayesian inference for marginal models under equality and inequality constraints</i>	599

Contributed Papers Session 12

Textual data

M. Dimai, N. Torelli	
<i>Clustering textual data by Latent Dirichlet Allocation: Applications and Extensions to Hierarchical Data</i>	605
D. F. Iezzi, M. Giorgino	
<i>Intimate femicide in Italy: a model to classify how killings happened</i>	609
S. Volo, C. Fisichella	
<i>Evaluating tourists experiences</i>	613

Contributed Papers Session 13

Data Analysis

M. Greenacre	
<i>Power Transformations in Correspondence Analysis and Related Methods</i>	619
G. Menardi, N. Torelli	
<i>Multimodal projection pursuit using the adjusted critical bandwidth</i>	623
R. Romano, T. Naes, P. Brockhoff	
<i>The use of analysis of variance and three-way factor analysis methods for studying the quality of a sensory panel</i>	627

POSTER SESSION

R. Belicchi, F. Crippa, L. Cesana, R. Daini	
<i>Lezak TinkerToy Test: freely making constructions to assess executive functions. An Italian normative sample</i>	633

A. Belliggiano, S. Staffieri	
<i>Private label organic food products: profile and behaviour of Italian consumers characterized through multivariate approach</i>	637
G. Boari, G. Cantaluppi, A. De Lauri	
<i>Handling missing values in PLS path modelling: comparison of alternative procedures</i>	641
S. Bozza, F. Taroni, F. Guy, M. Schmittbuhl	
<i>Taxonomic membership in great apes: a probabilistic approach from mandibular morphology</i>	645
S. De Cantis, M. Ferrante	
<i>Data quality and errors correction in Italian official statistics on guests of accommodation establishments</i>	649
C. Fukae	
<i>The Myths of Global Warming and Nuclear Power: Evidence from public opinion survey in Japan</i>	653
R. Gismondi, A. R. Giorgi, T. Pichiorri	
<i>Reducing Sampling Error using Deterministic Sample Selection in the Italian Retail Trade Monthly Survey</i>	657
R. Gismondi, M. A. Russo	
<i>Selection and Statistical Treatment of Variables for Evaluating Quality of Life in the Italian Provinces</i>	661
U. Magagnoli, P. M. Chiodini	
<i>Unilateral Hypothesis Tests of Efficiency Functions with Heteroscedastic Errors: an Iterative Procedure</i>	665
A. M. Olivieri	
<i>Excursionist flows and the attractiveness on the Aeolian Islands</i>	669
A. G. Quaranta, S. Fedeli, E. Voltattorni	
<i>Basel 2: firm's Scoring via Cluster Analysis and Artificial Neural Networks</i>	673

Ph.D. SPECIAL SESSION

G. Adelfio	
<i>Point processes residual analysis and asymptotical distribution of transformed versions of some second-order statistics</i>	679
V. Adorno	
<i>Program evaluation with continuous treatment: theoretical considerations and empirical application</i>	683

S. Bacci	
<i>Health related Quality of Life: some methodological aspects of Rasch analysis</i>	689
P. Eusebi	
<i>Many Facets Rasch Models: a Hierarchical Latent Class approach</i>	693
M. Matteucci	
<i>A multidimensional Item Response Theory approach for the University guidance</i>	697
R. Ragona	
<i>A State-Space Model for the Generation of Spatial Maps</i>	701
E. Romano	
<i>A new strategy for curves clustering based on free knots spline estimation</i>	705
I. Sulis	
<i>Measuring students' assessment on 'university course quality' using mixed-effects models</i>	709
Indice degli autori	713