



17th Conference of the
International Federation
of Classification Societies

Classification and Data Science in the Digital Age

Book of Abstracts



IFCS 2022

Book of Abstracts

17th Conference of the International Federation of
Classification Societies

Classification and Data Science
in the Digital Age

PORTO, PORTUGAL

Title: Classification and Data Science in the Digital Age - Book of Abstracts IFCS
2022

Authors: CLAD - Associação Portuguesa de Classificação e Análise de Dados

Cover design: exclamação!

Printed in Portugal by Instituto Nacional de Estatística

ISBN: 978-989-98955-9-1

N. DL: 500335/22

Number of copies: 350

Contents

Part I Presidential Address

Perturb and Conquer: how Classification Can Benefit from Data Perturbation	2
Angela Montanari	

Part II IFCS Research Medal Lecture

An Introduction to S-concordance and S-discordance	4
Edwin Diday	

Part III Tutorials

Analysis of Data Streams	6
João Gama	
Categorical Data Analysis and Visualization	7
Rosaria Lombardo and Eric J. Beh	

Part IV Keynote Lectures

A Showcase of New Methods for High Dimensional Data Viewing with Linear Projections and Sections	10
Dianne Cook	
Statistical Learning with Dynamic Interaction Data for Public Health . . .	11
Charles Bouveyron	
Fast Minipatch Ensemble Strategies for Learning and Inference	12
Genevera I. Allen	
Trends in Data Stream Mining	13
João Gama	

Part V Invited Sessions

A Criterion for Selecting the Number of Time Series Clusters	16
Daniel Peña and Ruey S. Tsay	
Clusters Based on Prediction Accuracy of Global Forecasting Models . . .	17
Pablo Montero-Manso, Ángel López-Oriona, and José A. Vilar	
Clustering and Classifying Time Series in the Sktime Toolkit: a Practical Review of Latest Advances in the Field	18
Anthony Bagnall	
Effect of Type 2 Diabetes and its Genetic Susceptibility on Severity and Mortality of COVID-19 in UK Biobank	19
Aeyeon Lee, Youngkwang Cho, Jun Li, Taesung Park, Wonil Chung, and Liming Liang	
Prognosis of COVID-19 Patients by the Underlying Diseases and Drug Treatment in Korea	20
Ho Kim and Taerim Lee	
Algorithms for Clustering COVID-19 Data: a Holistic Overview of Current Trends and New Visual Approaches	21
Eun-Kyung Lee	
Embedded Word MCA Biplots for Sentiment Visualisation: Application to COVID-19 Related Tweets	22
Zoë-Mae Adams, Johané Nienkemper-Swanepoel, Niël le Roux, and Sugnet Lubbe	
A General Framework for Implementing Distance Measures for Categorical Variables	23
Carlo Cavicchia, Michel van de Velden, Alfonso Iodice D’Enza, and Angelos Markos	
Some Descriptive Statistics of Aggregated Symbolic Data	24
Junji Nakano, Nobuo Shimizu, and Yoshikazu Yamamoto	
Variable Screening in High Dimensional Regression via Random Projection Ensembles	25
Laura Anderlucci, Matteo Farnè, Giuliano Galimberti, and Angela Montanari	
Model-based Clustering and Dimension Reduction for Multidimensional Social Networks	26
Michael Fop, Silvia D’Angelo, and Marco Alfò	
Conditional Gaussian Mixture Modeling	27
Volodymyr Melnykov and Yang Wang	

Classification Over Text, Relational Databases and Graphs - Software and Case Studies	28
Tomáš Kliegr	
Towards Deep and Interpretable Rule Learning	29
Johannes Fürnkranz	
Current Challenges in Interpretable Machine Learning and Partitioning Approaches	30
Bernd Bischl	
Part VI Benchmarking Challenge	
Vine Copula Mixture Models and Clustering for Non-Gaussian Data	32
Özge Sahin and Claudia Czado	
Comparison of Similarity Measures for Categorical Data in Hierarchical Clustering	33
Zdeněk Šulc and Hana Řezanková	
Comparing Model Selection Techniques to Determine the Number of Overlapping Clusters for the Additive Profile Clustering Model	34
Tom F. Wilderjans, Julian Rossbroich, and Jeffrey Durieux	
Pitfalls of Automatic Optimization Procedures and Benchmarking in Cluster Analysis	35
Quirin Stier and Michael C. Thrun	
Part VII Contributed Abstracts	
Biplots for Categorical and Ordinal Data Based on Logistic Responses . .	38
Jose Luis Vicente-Villardón	
The Biplot Inner Product for Interpretation and Derivation of Eigenvector Methods	39
Cajo J. F. ter Braak	
Fifty Years of Biplots: Some Remaining Enigmas and Challenges	40
Jan Graffelman	
Outlier Detection for BIG Functional Data	41
Rosa E. Lillo, Oluwasegun T. Ojo, and Antonio Fernández-Anta	
Outlier and Novelty Detection for Functional Data: a Semiparametric Bayesian Approach	42
Francesco Denti, Andrea Capozzo, and Francesca Greselin	
A Geometric Perspective on Functional Outlier Detection	43
Moritz Herrmann and Fabian Scheipl	

Contents	xix
A New Decomposition of Orthogonal Matrices with Application to Common Principal Components	44
Luca Bagnato and Antonio Punzo	
An MML Embedded Approach for Estimating the Number of Clusters . .	45
Cláudia Silvestre, Margarida G. M. S. Cardoso, and Mário Figueiredo	
Comparison of Segmentation Approaches for Partial Least Squares Path Modeling with Stability Assessment	46
Sophie Dominique, Mohamed Hanafi, Fabien Llobell, Jean-Marc Ferrandi, and Véronique Cariou	
Robust Classification for Toroidal Data	47
Giovanni Saraceno, Luca Greco, and Claudio Agostinelli	
Consistency of Trimmed Estimators of Scatter Under the t-distribution . .	48
Andrea Cerioli, Lucio Barabesi, Luis A. García-Escudero, and Agustín Mayo-Iscar	
Robust Classification in High Dimensions Using Regularized Covariance Estimates	49
Valentin Todorov and Peter Filzmoser	
Symbolic Concordance and Discordance Illustrated on Data from an International Teaching and Learning Survey	50
Simona Korenjak-Černe, Barbara Japelj Pavešić, and Edwin Diday	
A Clusterwise Regression Method for Distributional Data	51
Rosanna Verde, Antonio Balzanella, and Antonio Irpino	
The Use of Regression to Partition a Dataset of Interval Observations . . .	52
Lynne Billard and Fei Liu	
Heterogeneous Random Forests	53
Ye-eun Kim and Hyunjoong Kim	
Analysis of the Damage Rate Using Typhoon Information	54
Su Hoon Choi and Min Soo Kim	
Resampling, Relabeling, and Raking for Extremely Imbalanced Classification	55
Hae-Hwan Lee, Seunghwan Park, and Jongho Im	
Clustering Student Mobility Data in 3-way Networks	56
Vincenzo Giuseppe Genova, Giuseppe Giordano, Giancarlo Ragozini, and Maria Prosperina Vitale	
Multi-perspective Risky User Classification in Social Networks	57
Antonio Pellicani, Gianvito Pio, and Michelangelo Ceci	

Clustering and Blockmodeling Temporal Networks – Two Indirect Approaches	58
Vladimir Batagelj	
Clustering Validation in Hierarchical Cluster Analysis: an Empirical Study	59
Osvaldo Silva, Áurea Sousa, and Helena Bacelar-Nicolau	
Divide and Conquer: a Clustering Method for Hierarchical and Nested Data Structures	60
Andrej Svetlošák, Miguel de Carvalho, Gabriel Martos Venturini, and Raffaella Calabrese	
Significance Mode Analysis (SigMA) for Hierarchical Structures	61
Sebastian Ratzenböck, Torsten Möller, Josefa E. Großschedl, João Alves, Immanuel M. Bomze, and Stefan Meingast	
Kurtosis-based Projection Pursuit for Matrix-valued Data	62
Una Radojicic, Klaus Nordhausen, and Joni Virta	
Comparison of Pixel Based Segmentation Methods in Papillary Thyroid US Images	63
Neslihan Gökmen İnan, İsmail Meşe, Düzgün Yıldırım, and Ozan Kocadağlı	
Bootstrapping Binary GEV Regressions for Massive Unbalanced Datasets	64
Michele La Rocca, Marcella Niglio, and Marialuisa Restaino	
Stochastic Collapsed Variational Inference for Structured Gaussian Process Regression Networks	65
Rui Meng, Herbert K. H. Lee, and Kristofer Bouchard	
Covariate Selection Method in Propensity Score Model for the Quantile Treatment Effect Estimation	66
Takehiro Shoji, Jun Tsuchida, and Hiroshi Yadohisa	
Are Attitudes Toward Immigration Changing in Europe? An Analysis Based on Latent Class IRT Models	67
Ewa Genge and Francesco Bartolucci	
Visualization of IATA Regions in Air Transport Before and After the COVID-19 Pandemic	68
Tüzün Tolga İnan, Neslihan Gökmen İnan, Aylin Yaman Kocadağlı, and Ozan Kocadağlı	
Political and Religion Attitudes in Greece: Behavioral Discourses	69
Georgia Panagiotidou and Theodore Chadjipadelis	
Functional Data Representation with Merge Trees	70
Matteo Pegoraro and Piercesare Secchi	

Contents	xxi
Elastic Regression for Irregularly Sampled Curves in \mathbb{R}^d	71
Lisa Steyer, Almond Stöcker, and Sonja Greven	
Misalignment of Spectral Data: Constrained Optimization in a Functional Data Analysis Framework	72
Francesca Di Salvo, Delia Francesca Chillura Martino, and Gabriella Chirco	
Model Based Clustering of Functional Data with Mild Outliers	73
Cristina Anton and Iain Smith	
Old and New Constraints in Model Based Clustering	74
Luis A. García-Escudero, Agustín Mayo-Iscar, Gianluca Morelli, and Marco Riani	
Model Based Clustering and Outlier Detection with Missing Data	75
Cristina Tortora, Hung Tong, and Louis Tran	
How to Mitigate the Effect of Outliers on Balancing Technique	76
Rasool Taban, Maria do Rosário Oliveira, and Claudia Nunes Philippart	
Outliers Detection in Functional Data	77
Amovin-Assagba Martial, Gannaz Irène, and Jacques Julien	
Robustified Elastic Net Estimator for Multinomial Regression	78
Fatma Sevinç Kurnaz and Peter Filzmoser	
Optimized Symbolic Correspondence Analysis for Multi-valued Variables	79
Jorge Arce Garro and Oldemar Rodríguez Rojas	
Symbolic t-SNE and UMAP Methods for Interval Type Variables.	80
Oldemar Rodríguez Rojas	
Two-stage Principal Component Analysis on Interval-valued Data Using Patterned Covariance Structure	81
Anuradha Roy	
Detection of the Biliary Atresia Using Deep Convolutional Neural Networks Based on Statistical Learning Weights via Optimal Similarity and Resampling Methods	82
Kuniyoshi Hayashi, Eri Hoshino, Mitsuyoshi Suzuki, Erika Nakanishi, Kotomi Sakai, and Masayuki Obatake	
Variational Autoencoder with Gamma Mixture for Clustering Right-skewed Data	83
Jinwon Heo and Jangsun Baek	
An Efficient Way to Identify Inliers via Inlier-memorization Effect of Deep Generative Models	84
Dongha Kim, Jaesung Hwang, and Yongdai Kim	

Three-way Spectral Clustering	85
Cinzia Di Nuzzo and Salvatore Ingrassia	
Fuzzy Clustering by Hyperbolic Smoothing	86
David Masís, Esteban Segura, Javier Trejos, and Adilson Xavier	
Combining KDE and DBSCAN Clustering to Understand Road Traffic Accidents: the Case of Setúbal, Portugal	87
Pedro Nogueira, Marcelo Silva, Paulo Infante, Paulo Rebelo Manuel, Leonor Rego, Anabela Afonso, and Gonçalo Jacinto	
Similarity Forest for Time Series Classification	88
Tomasz Górecki, Maciej Łuczak, and Paweł Piasecki	
Uncovering Regions of Maximum Dissimilarity on Random Process Data	89
Miguel de Carvalho and Gabriel Martos Venturini	
Franz Liszt's Transcendental Études: an Evolutionary Analysis by Machine Learning	90
Matteo Farnè	
Quantile-distribution Functions and Their Use for Classification	91
Eduardo Redivo, Cinzia Viroli, and Alessio Farcomeni	
Analysis of Gini Splitting Criterion and Comparison with Maximum Likelihood Rule	92
Amirah S. Alharthi and Charles C. Taylor	
Envelope-based Support Vector Machine Classifier	93
Alya Alzahrani and Andreas Artemiou	
A Moment-free Measure of Multivariate Skewness	94
Andrzej Sokolowski and Malgorzata Markowska	
The Weighted RV Coefficient: Exact Moments by Invariant Orthogonal Integration	95
François Bavaud	
Testing Equality of Multivariate Coefficients of Variation	96
Marc Ditzhaus and Łukasz Smaga	
A New Regression Model for the Analysis of Microbiome Data	97
Roberto Ascari and Sonia Migliorati	
The Death Process in Italy Before and During the Covid-19 Pandemic: a Functional Compositional Approach	98
Riccardo Scimone, Alessandra Menafoglio, Laura M. Sangalli, and Piercesare Secchi	

Sampling Design for Uncovering Natural Laws in Compositional Data . . .	99
Lan Liang, Glòria Mateu-Figueras, and Jan Graffelman	
Penalized Model-based Functional Clustering: a Regularization Approach via Shrinkage Methods	100
Nicola Pronello, Rosaria Ignaccolo, Luigi Ippoliti, and Sara Fontanella	
Clustering in FDA Mixing the Epigraph and the Hypograph Indexes with Machine Learning Algorithms	101
Belén Pulido, Alba M. Franco-Pereira, and Rosa E. Lillo	
Localization Processes for Functional Data Classification	102
Antonio Elías, Raúl Jiménez, and Joseph E. Yukich	
A New Functional Data Clustering Technique Based on Spectral Clustering and Downsampling	103
Maryam Al Alawi, Surajit Ray, and Mayetri Gupta	
Parsimonious Mixtures of Seemingly Unrelated Contaminated Normal Regression Models	104
Gabriele Perrone and Gabriele Soffritti	
Monitoring Hyperparameter Choice for Robust Cluster Weighted Model	105
Andrea Cappozzo, Luis A. García-Escudero, Francesca Greselin, and Agustín Mayo-Iscar	
Latent Block Regression Model	106
Rafika Boutalbi, Lazhar Labiod, and Mohamed Nadif	
Towards a Bi-stochastic Matrix Approximation of k-means and Some Variants	107
Lazhar Labiod and Mohamed Nadif	
Clustering Brain Connectomes Through a Density-peak Approach	108
Riccardo Giubilei	
New Metrics for Classifying Phylogenetic Trees Using k-means and the Symmetric Difference Metric	109
Nadia Tahiri and Aleksandr Koshkarov	
Alternating Optimization Framework for Sparse Simultaneous Component Analysis Based on Data Integration	110
Rosember Guerra-Urzola, Juan C. Vera, Katrijn Van Deun, and Klaas Sijtsma	
Joint Sparse Principal Component Analysis	111
Katrijn Van Deun	
Joint Sparse Principal Component Analysis: a Simulation Study	112
Tra Le and Katrijn Van Deun	

Copula-based Non-metric Unfolding on Augmented Data Matrix	113
Marta Nai Ruscone and Antonio D'Ambrosio	
Emotion Classification Based on Single Electrode Brain Data: Applications for Assistive Technology	114
Duarte Rodrigues, Luis Paulo Reis, and Brígida Mónica Faria	
On the Role of Data, Statistics and Decisions in a Pandemic	115
Ursula Garczarek, Beate Jahn, Sarah Friedrich, Joachim Behnke, Joachim Engel, Ralf Münnich, Markus Pauly, Adalbert Wilhelm, Olaf Wolkenhauer, Markus Zwick, Uwe Sieber, and Tim Friede	
A Deep Learning Analytics to Detect Dental Caries	116
Taerim Lee	
Identification of Shared Genetic Loci Between Psychiatric Disorders and Telomere Length and Evaluation of Their Role as Potential Drug Targets	117
Claudia Pisanu, Anna Meloni, and Alessio Squassina	
Estimating Optimal Decision Trees for Treatment Assignment with $k > 2$ Treatment Alternatives: a Classification Problem with a Unit- and Class- dependent Misclassification Cost	118
Iven Van Mechelen and Aniek Sies	
ExactTree: an R-package for Globally Optimal Decision Trees	119
Elise Dusseldorp, Juan Claramunt Gonzales, Jacqueline Meulman, Samil Uysal, and Bart Jan van Os	
Optimal Random Projection Trees Ensemble	120
Nosheen Faiz, Adi Lausen, Metodi Metodiev, Zardad Khan, and Berthold Lausen	
Born-again and Bayesian Approaches for Improving the Performance of Decision Trees	121
Marjolein Fokkema	
Evolution of Media Coverage on Climate Change and Environmental Awareness: an Analysis of Tweets from UK and US Newspapers	122
Gianpaolo Zammarchi, Maurizio Romano, and Claudio Conversano	
Improving Classification of Documents by Semi-supervised Clustering in a Semantic Space	123
Jasminka Dobša and Henk A.L. Kiers	
Is It Hate or Criticism? An Exploratory Approach to Negative Comments on YouTube	124
Manuela Schmidt	

A Time-varying Text Based Ideal Point Model to Infer Partisanship in the U.S. Senate	125
Sourav Adhikari, Bettina Grün, and Paul Hofmarcher	
Oracle-LSTM: a Neural Network Approach to Mixed Frequency Time Series Prediction	126
Alessandro Bitetto and Paola Cerchiello	
Time Series of Counts Under Censoring	127
Isabel Silva, Maria Eduarda Silva, Isabel Pereira, and Brendan McCabe	
Multivariate Time Series Feature Extraction via Multilayer Networks . . .	128
Vanessa Freitas Silva, Maria Eduarda Silva, Pedro Ribeiro, and Fernando Silva	
On the Use of the Choquet Fuzzy Integral to Aggregate Predictions of Time Series: an Application to Economic (and Other Types of) Data	129
Diogo Alves, José Matos, and Sandra Silva	
Some Biplot Alternatives	130
Patrick Groenen	
Biplots in Dimension Reduction and Clustering	131
Alfonso Iodice D’Enza, Angelos Markos, and Michel van de Velden	
Biplots: a Sophisticated Multivariate Approach or a User-Friendly Technique?	132
Manuel Rui Alves	
PLS-based Principal Balances for Regression and Classification with High-dimensional Compositional Data	133
Viktorie Nesrstová, Ines Wilms, Karel Hron, Josep A. Martín-Fernández, Peter Filzmoser, and Javier Palarea-Albaladejo	
Clustering Count Data Using Compositional Methods	134
Marc Comas-Cufí, Josep A. Martín-Fernández, Glària Mateu-Figueras, and Javier Palarea-Albaladejo	
Urban Development Paths in Poland: Multidimensional Perspective	135
Barbara Batóg and Jacek Batóg	
Analyzing the Evolution of EU Countries and Indicators of Europe 2020 Agenda	136
Adelaide Figueiredo and Fernanda Figueiredo	
Google Trends as a Macroeconomic Predictor: Behind the Scenes	137
Eduardo Andre Costa and Maria Eduarda Silva	

COVID-19 Pandemic: a Methodological Model for the Analysis of Government Preventing Measures and Health Data Records	138
Theodore Chadjipadelis and Sofia Magopoulou	
Detecting Fabricated Interviews Using the Hamming Distance	139
Joerg Blasius	
Digital Development and Internet Use in the European Union Countries .	140
Fernanda Figueiredo and Adelaide Figueiredo	
Probabilistic Clustering with Local Aligement of Italian COVID-19 Death Curves	141
Marzia A. Cremona, Tobia Boschi, and Francesca Chiaromonte	
Model Free Predictive Inference for Functional Kriging Techniques Based on Conformal Prediction	142
Andrea Diana, Elvira Romano, and Jorge Mateu	
Density Modelling via Functional Data Analysis	143
Stefano A. Gattone and Tonio Di Battista	
On Parsimonious Modelling via Matrix-variate t Mixtures	144
Salvatore D. Tomarchio	
Four Skewed Tensor Variate Distributions	145
Michael P.B. Gallagher, Peter A. Tait, and Paul D. McNicholas	
A Family of Skewed Power Exponential Mixture Models for Clustering and Classification	146
Utkarsh J. Dang, Michael P. B. Gallagher, Ryan P. Browne, and Paul D. McNicholas	
Generating Collective Counterfactual Explanations in Score-based Classification via Mathematical Optimization	147
Jasone Ramírez-Ayerbe, Emilio Carrizosa, and Dolores Romero Morales	
Spherical Separation in Machine Learning	148
Matteo Avolio, Annabella Astorino, and Antonio Fuduli	
Model Extraction Based on Counterfactual Explanations	149
Cecilia Salvatore and Veronica Piccialli	
Isolation Forests for Symbolic Data as a Tool for Outlier Mining	150
Andrzej Dudek and Marcin Pelka	
Symbolic Clustering Methods Applied to Interval Estimates of Production Cost Quantiles	151
Dominique Desbois	

Fisher Discriminant Analysis for Interval Data	152
Diogo Pinheiro, Maria do Rosário Oliveira, Igor Kravchenko, and Lina Oliveira	
Stability of Mixed-type Cluster Partitions for Determination of the Number of Clusters	153
Rabea Aschenbruck, Gero Szepannek, and Adalbert Wilhelm	
Multinomial Multilevel Models with Discrete Random Effects: a Multivariate Clustering Tool	154
Chiara Masci, Francesca Ieva, and Anna Maria Paganoni	
PD-clustering for Mixed Data Type	155
Francesco Palumbo and Cristina Tortora	
Anomaly Detection-based Under-sampling for Imbalanced Classification Problems	156
You-Jin Park, Chun-Yang Peng, Rong Pan, and Douglas C. Montgomery	
Continuous Adaptation to Distribution Drifts Through Continual Learning in Manufacturing	157
Henrique Siqueira and Onay Urfalioglu	
Detecting Anomalies with TADGAN: a Case Study	158
Inês Oliveira e Silva, Carlos Soares, Arlete Rodrigues, and Pedro Bastardo	
A Trivariate Geometric Classification of Decision Boundaries for Mixtures of Regressions	159
Filippo Antonazzo and Salvatore Ingrassia	
Clustering Rainfall by Simulated Annealing for Histogram Symbolic Data	160
Alejandro Chacón and Javier Trejos	
Statistical Assessment of Youth Inclusion in the National Labour Markets	161
Beata Bal-Domańska	
Barriers to Industry Digitization in Poland from the Perspective of High and Medium-high Technology Sector Enterprises	162
Elżbieta Sobczak, Marcin Pelka, and Karolina Pokorska	
Kernel Smoothing-based Probability Contours for Tumour Segmentation	163
Wenhui Zhang and Surajit Ray	
Parameter Estimation for Mixtures of Linear Mixed Models: the EM, CEM and SEM Algorithms	164
Luísa Novais and Susana Faria	

Genomic Prediction Using Machine Learning Methods: Performance Comparison on Synthetic and Empirical Data	165
Vanda Lourenço, Joseph Ogutu, Rui Rodrigues, and Hans-Peter Piepho	
Pooled Mean and Confidence Interval Estimation Combining Different Sets of Summary Statistics	166
Flora Ferreira, José Soares, Fernanda Sousa, Filipe Magalhães, Isabel Ribeiro, and Pedro Pacheco	
Prediction of Diabetes via Bayesian Network Classifier from Exposure to Environmental Polluting Chemicals Data	167
Rosy Oh, Hong Kyu Lee, Youngmi Kim Pak, and Man-Suk Oh	
Model Performance Metrics for Sample Selection Bias Correction by Pseudo Weighting	168
An-Chiao Liu, Ton de Waal, Katrijn Van Deun, and Sander Scholtus	
Some Factors that Influence the Nature of Road Traffic Accidents	169
Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Leonor Rego, Vitor Nogueira, Paulo Quaresma, José Saias, Daniel Santos, Pedro Nogueira, Marcelo Silva, Rosalina Pisco Costa, Patrícia Gois, and Paulo Rebelo Manuel	
Comparing Variable Selection Methods for High-dimensional Compositional Data in a Discriminant Analysis Context	170
Pepus Daunis-i-Estadella, Glòria Mateu-Figueras, Viktorie Nesrstová, Karel Hron, and Josep A. Martín-Fernández	
Cluster Analysis and Genetic Risk Score in Age-related Macular Degeneration - the Coimbra Eye Study	171
Rita Coimbra	
Sensor System for Standardizing Articulation Patterns According to Korean Phonemes	172
Seong Tak Woo and Da Hee Oh	
Categorical Data Visualization and the Cressie-Read Divergence Statistic	173
Eric J. Beh and Rosaria Lombardo	
Biplots Based on Latent Variable Models in the Analysis of Ecological Communities	174
Jenni Niku and Sara Taskinen	
Biplot Representation of Partial Least Squares Regression for Binary Responses	175
Laura Vicente-Gonzalez and Jose Luis Vicente-Villardón	
Generalized Spatio-temporal Regression with PDE Penalization	176
Eleonora Arnone, Elia Cunial, and Laura M. Sangalli	

Impact Point Selection in Semiparametric Bifunctional Models	177
Silvia Novo, Germán Aneiros, and Philippe Vieu	
Latent Function-on-scalar Regression Models for Observed Sequences of Correlated Binary Data: a Restricted Likelihood Approach	178
Fatemeh Asgari and Valeria Vitelli	
pcTVI: Parallel MDP Solver Using a Decomposition into Independent Chains	179
Jaël Champagne Gareau, Éric Beaudry, and Vladimir Makarenkov	
Classification of Viral Pneumonia Images via Multiple Instance Learning	180
Antonio Fuduli, Matteo Avolio, Eugenio Vocaturo, and Ester Zumpano	
Nonlinear Approaches for Multiple Instance Learning	181
Annabella Astorino, Matteo Avolio, and Antonio Fuduli	
An Online Minorization-Maximization Algorithm	182
Hien Duy Nguyen, Florence Forbes, Gersende Fort, and Olivier Cappé	
Frugal Gaussian Clustering of Huge Ombalanced Datasets Through a Bin-marginal Approach	183
Filippo Antonazzo, Christophe Biernacki, and Christine Keribin	
Reinforced EM Algorithm Through Clever Initialization for Clustering with Gaussian Mixture Models	184
Joshua Tobin, Chin Pang Ho, and Mimi Zhang	
Exact Computation of the Angular Halfspace Depth	185
Stanislav Nagy and Rainer Dyckerhoff	
Reconstruction of Atomic Measure Based on its Simplicial Depth	186
Petra Laketa and Stanislav Nagy	
Robustness Aspects of Optimized Centroids	187
Jan Kalina and Patrik Janáček	
Analysis of the Changes in the Polish Traditional Drugstores Market During COVID-19	188
Marcin Pelka, Antonio Irpino, and Michal Swachta	
Logistic Regression Models for Aggregated Data	189
Thomas Whitaker, Boris Beranger, and Scott Sisson	
Nonparametric Regressions for Distributional Data	190
Albert Meco, Javier Arroyo, and Antonio Irpino	

Hotspot Cluster Detection Based on Spatial Hierarchical Structure and its Software	191
Fumio Ishioka, Shoji Kajinishi, and Koji Kurihara	
Group Lasso Penalty for Spatially Clustered Coefficient Regression	192
Toshiki Sakai, Jun Tsuchida, and Hiroshi Yadohisa	
Visualization of the Number of New Positives for COVID-19 in Japan	193
Yoshiro Yamamoto, Sanetoshi Yamada, Mayumi Tanahashi, and Tadashi Imanishi	
Two Simple but Efficient Algorithms to Recognize Robinson Dissimilarities	194
Mikhaël Carmona, Guylain Naves, Victor Chepoi, and Pascal Pr�ea	
Clustering with Missing Data: Which Imputation Model for Which Cluster Analysis Method?	195
Vincent Audigier, Nd�ye Niang, and Matthieu Resche-Rigon	
Hierarchies and Weak-hierarchies as Interval Convexities	196
Patrice Bertrand and Jean Diatta	
A Rule-based Approach to Scoring Systems	197
Michael Rapp, Johannes F�rnkranz, and Eyke H�llermeier	
On Explaining Model Change Based on Feature Importance	198
Maximilian Muschalik, Fabian Fumagalli, Barbara Hammer, and Eyke H�llermeier	
Interpretable Multi-class Trees for Travel Choice Mode Analysis	199
Christian Riccio, Andrea Papola, Michele Staiano, and Roberta Siciliano	
Identification of Driver Genes in Glioblastoma via Regularized Classification	200
Marta Belchior Lopes and Susana Vinga	
Outlier Detection: a Procedure to Capture Atypical Groups of Observations	201
Ana Helena Tavares, Vera Afreixo, and Paula Brito	
Bayesian Classification and Non-Bayesian Label Estimation via EM Algorithm to Identify Differential Expression in Omics Data: a Comparative Study	202
Marilia Antunes and Lisete Sousa	
Sequence-aware Item Recommendations for Multiply Repeated User-item Interactions	203
Juan Pablo Equihua, Maged Ali, Henrik Nordmark, and Berthold Lausen	

Contents	xxxi
High-dimensional Linear Regression Estimation	204
Mauro Iannuzzi and Matteo Farnè	
Experimental Study of Similarity Measures for Clustering Uncertain Time Series	205
Michael Dinzinger, Michael Franklin Mbouopda, and Engelbert Mephu Nguifo	
Assessing the Status of Two Data-limited Skates Landed in Portuguese Ports Using an Empirical Catch Rule	206
Erick Chatalov, Ivone Figueiredo, Lisete Sousa, and Bárbara Pereira	
Machine Learning Approach to Identify Factors that Influence Accident Severity	207
Daniel Santos, Vitor Nogueira, José Saias, Paulo Quaresma, Paulo Infante, Gonçalo Jacinto, Anabela Afonso, Leonor Rego, Pedro Nogueira, Marcelo Silva, Rosalina Pisco Costa, Patrícia Gois, and Paulo Rebelo Manuel	
Trade and Bank Credit of Portuguese SMEs: a Panel Data Application . .	208
Carla Henriques, Pedro Pinto, and Carolina Cardoso	
Hausdorff Distance: a Powerful Tool for Matching Households and Individuals in Historical Censuses	209
Thais Pacheco Menezes, Michael Fop, and Thomas Brendan Murphy	
Model-based Tri-clustering	210
Rafika Boutalbi, Lazhar Labiod, and Mohamed Nadif	
Analyzing the Effects of Deviations from Normality on the Latent Growth Curve Models Goodness-of-fit	211
Catarina Marques, Maria de Fátima Salgueiro, and Paula C.R. Vicente	
Transformation Mixture Modeling for Skewed Data Groups with Heavy Tails	212
Yana Melnykov, Xuwen Zhu, and Volodymyr Melnykov	
A Simulation Study on Variable Selection in Mixture Regression Models .	213
Susana Faria	
Hybrid Forecasting Combinations by Feature Based Metalearning	214
Moises Santo, Andre C.P.L.F. de Carvalho, and Carlos Soares	
On the Measurement of Household Subjective Poverty: Concepts and Application	215
Aleksandra Łuczak and Sławomir Kalinowski	

Socio-economic Classification of Territorial Units: Extreme Value Theory-based Methods as Support for the Construction of a Synthetic Index	216
Aleksandra Łuczak and Małgorzata Just	
Depth-based Two-sample Testing	217
Felix Gnetner, Claudia Kirch, and Alicia Nieto-Reyes	
The Control of False Discovery Rate for Functional Data	218
Niels Lundtorp Olsen, Alessia Pini, and Simone Vantini	
Functional Random Forest for Biomedical Signals Classification and Interpretative Tools	219
Fabrizio Maturò and Rosanna Verde	
Correlation-based Iterative Clustering Methods for Time Course Data . .	220
Michelle Carey, Shuang Wu, Guojun Gan, and Hulin Wu	
Depth-based Classifiers for Partially Observed Functional Data	221
Antonio Elías, Raúl Jiménez, Anna Maria Paganoni, and Laura M. Sangalli	
Using Clustering and Machine Learning Methods to Provide Intelligent Grocery Shopping Recommendations	222
Nail Chabane, Mohamed Achraf Bouaoune, Reda Amir Sofiane Tighilt, Bogdan Mazouze, Nadia Tahiri, and Vladimir Makarenkov	
Typology of Motivation Factors for Employees in the Banking Sector: Multivariate Data Analysis	223
Áurea Sousa, Osvaldo Silva, M. Graça Batista, Sara Cabral, and Helena Bacelar-Nicolau	
Industry Sector Detection in Legal Articles Using Transformer-based Deep Learning	224
Hui Yang, Stella Hadjiantoni, Yunfei Long, Rūta Petraitytė, and Berthold Lausen	
User Segmentation Based on Online Behavioural Data via Ensemble Predictions and Clustering	225
Stella Hadjiantoni, Hui Yang, Yunfei Long, Ruta Petraityte, and Berthold Lausen	
Attitudes Toward Statistics in the 3rd Cycle of Basic Education in Portugal	226
Adelaide Freitas, Ana Julieta Morais, Pedro Sá Couto, and Anabela Rocha	
Predictors of Quantitative Skills in Degree Schemes at University	227
Alex Partner, Adi Lausen, Alexei Vernitski, Chris Saker, and Berthold Lausen	

Using Excel and R for Teaching Statistics and Data Analysis	228
W.H. Moolman	
Students' Assessment Through a IRT and Archetypal Analysis Joint Strategy	229
Lucio Palazzo and Francesco Palumbo	
Kernel-based Hierarchical Structural Component Models for Pathway Analysis	230
Suhyun Hwangbo, Sungyoung Lee, Seungyeoun Lee, Heungsun Hwang, Inyoung Kim, and Taesung Park	
Bayesian Inference for the Generation Interval of COVID-19 in Busan, Korea	231
Jayoeng Paek, Ilsa Choi, Kyeongah Nah, and Yongkuk Kim	
Fitting an Accelerated Failure Time Model with Time-dependent Covariates via Nonparametric Gaussian Scale Mixtures	232
Ju-Young Park, Byungtae Seo, and Sangwook Kang	
Comparison of Survival Prediction Models for Pancreatic Cancer: Cox Model vs. Machine Learning Models	233
Hyunsuk Kim, Taesung Park, and Seungyeoun Lee	
Clustering High-dimensional Microbiome Data	234
Sanjeena Dang (Subedi) and Wangshu Tu	
Clustering Adolescent Female Physical Activity Levels with an Infinite Mixture Model on Random Effects	235
Amy LaLonde, Tanzy Love, Deborah Rohm Young, and Tongtong Wu	
Modeling Three-way RNA Sequencing Data Using Data Transformations and Matrix-variate Gaussian Mixture Models	236
Theresa Scharl and Bettina Grün	
Some Issues in Robust Clustering	237
Christian Hennig	
Assessing Common Principal Directions	238
David Rodríguez Vítóres and Carlos Matrán Bea	
Robustness and Initialization Issues in Subspace Clustering	239
Luis A. García-Escudero and Agustín Mayo-Iscar	
A Likelihood Ratio Test for Choosing Input Parameters in Robust Model Based Clustering	240
Luis A. García-Escudero, Agustín Mayo-Iscar, Gianluca Morelli, and Marco Riani	

Data Clustering and Representation Learning Based on Networked Data	241
Lazhar Labiod and Mohamed Nadif	
Exploratory Graph Analysis for Configural Invariance Assessment of a Test	242
Alex Cucco, Lara Fontanella, Sara Fontanella, and Nicola Pronello	
An Extension of Edge Reduction for Large Networks	243
Pedro Campos	
Patterns of Cooperation for Polish Authors of Research Publications in Economics, Business and Medicine Areas	244
Urszula Cieraszewska, Paweł Lula, Magdalena Talaga, and Marcela Zembura	
A Topological Clustering of Individuals	245
Rafik Abdesselam	
Modeling a Most Specific Generalization in Domain Taxonomies	246
Zhirayr Hayrapetyan, Boris Mirkin, Susana Nascimento, Trevor Fenner, and Dmitry Frotov	
A Proposal for Formalization and Definition of Anomalies in Dynamical Systems	247
Jan Michael Spoor, Jens Weber, and Jivka Ovtcharova	
Unsupervised Classification of Categorical Time Series Through Innovative Distances	248
Ángel López-Oriona, José A. Vilar, and Pierpaolo D'Urso	
Detecting Differences in Italian Regional Health Services During Two Covid-19 Waves	249
Lucio Palazzo and Riccardo Ievoli	
The Clustering Performance of a Weighted Combined Distance Between Time Series	250
Margarida G. M. S. Cardoso, Ana Alexandra Martins, and João Lagarto	
Dimensionality Reduction and Multivariate Time Series Classification	251
Veronne Yepmo, Angeline Plaud, and Engelbert Mephu Nguifo	
A Review on Official Survey Item Classification for Mixed-Mode Effects Adjustment	252
Afshin Ashofteh and Pedro Campos	
Adaptive Fuzzy Systems in Economics and Finance: Evaluating Interval Forecasts of High-frequency Data	253
Rosangela Ballini	

The Usefulness of Selected Machine Learning Methods for Estimating Missing Data to Supplement Databases Used for Corporate Bankruptcy Prediction	254
Barbara Pawelek and Jozef Pocięcha	
Registration of 24-hour Accelerometric Rest-activity Profiles and its Application to Human Chronotypes	255
Erin I. McDonnell, Vadim Zipunnikov, Jennifer A. Schrack, Jeff Goldsmith, and Julia Wrobel	
Functional Data from Wearable Devices: a Review	256
Nihan Acar-Denizli and Pedro Delicado	
A Wavelet-mixed Effect Landmark Model for the Effect of Potassium and Biomarkers Profiles on Survival in Heart Failure Patients	257
Caterina Gregorio, Giulia Barbati, and Francesca Ieva	
True Sparsity Approaches in Classification via Conic Optimization	258
Immanuel M. Bomze and Bo Peng	
Creating Homogeneous Sectors: Criteria and Applications of Sectorization	259
Cristina Lopes, Maria Margarida Lima, Elif Göksu Öztürk, Ana Maria Rodrigues, Ana Catarina Nunes, Cristina Oliveira, José Soeiro Ferreira, and Pedro Filipe Rocha	
MARGOT: a Maximum MARGIn Optimal Classification Tree	260
Federico D’Onofrio, Marta Monaci, Giorgio Grani, and Laura Palagi	
Multivariate Mapping of Soil Organic Carbon and Nitrogen	261
Stephan van der Westhuizen, David P. Hofmeyr, and Gerard B.M. Heuvelink	
Spatial Configuration of Fire Stations in Portugal	262
Regina Bispo, Clara Yokochi, Francisca G. Vieira, Nádia Bachir, Pedro Espadinha-Cruz, José Pedro Lopes, Alexandre Penha, Marta Belchior Lopes, Filipe J. Marques, João Paulo Rodrigues, and António Grilo	
Spatio-temporal Variability of Distribution and Abundance of Sardine in Portuguese Continental Coast: Environmental Effects	263
Daniela Silva, Raquel Menezes, Ana Moreno, Ana Teles-Machado, and Susana Garrido	
Time Resolved Feature Importance of a Biopharmaceutical Purification Process Using Permutation Based Methods	264
Matthias Medl, Theresa Scharl, Astrid Dürauer, and Friedrich Leisch	

Off-target Predictions in CRISPR-Cas9 Gene Editing Using Machine Learning	265
Ali Mertcan Kose and Ozan Kocadağlı	
Comparison of k-mer and Alignment-based Pre-processing Approaches for Machine Learning Based Functional Annotation with 16S rRNA Data	266
Rafal Kulakowski, Adi Lausen, Etienne Low-Decarie, and Berthold Lausen	
An Ultrametric Model for Clustering and Dimensionality Reduction	267
Giorgia Zaccaria	
Combining Latent Class Analysis and Multiple Correspondence Analysis	268
Alice Barth	
Simultaneous Factorial Reduction and Clustering for Three-mode Data Sets: a Comparison	269
Prosper Ablordeppy, Adelaide Freitas, Maurizio Vichi, and Giorgia Zaccaria	
Clustering Intensive Longitudinal Data Through Mixture Multilevel Vector-autoregressive Modeling	270
Anja Ernst, Marieke Timmerman, Feng Ji, Bertus Jeronimus, and Casper Albers	
Misppecification Tests for Hidden Markov Models Based on a New Class of Finite Mixture Models	271
Francesco Bartolucci, Silvia Pandolfi, and Fulvia Pennoni	
Natural Cubic Smoothing Splines for Latent Class Identification in Longitudinal Growth Trajectories	272
Katerina M. Marcoulides and Laura Trinchera	
Supervised Classification via Neural Networks for Replicated Point Patterns	273
Kateřina Pawlasová, Iva Karafiátová, and Jiří Dvořák	
Reliability Assessment of Ancient Stone Arch Bridge Applying ANN models, Case Study: Leça Railway Bridge	274
Edward A. Baron, Ana Margarida Bento, José Campos e Matos, Rui Calçada, and Kenneth Gavin	
Application of Artificial Intelligence (AI) in Flood Risk Forecasting	275
Minh Quang Tran, Ana Margarida Bento, Elisabete Teixeira, Helder Sousa, and José Campos e Matos	
Logistic Regression with Sparse Common and Distinctive Covariates	276
Soogeun Park, Eva Ceulemans, and Katrijn Van Deun	

Accuracy Measures for Binary Classification Based on Quantitative Group Tests	277
Rui Santos, João Paulo Martins, and Miguel Felgueiras	
Exploiting Pareto Density Estimation for Nonparametric Naïve Bayes Classifiers	278
Quirin Stier and Michael C. Thrun	
Author Index	279

Outlier and Novelty Detection for Functional Data: a Semiparametric Bayesian Approach

Francesco Denti, Andrea Cappelozzo, and Francesca Greselin

A novelty detection model can be seen as a supervised classifier, trained on a fully-labeled training set, that allows for the presence of new classes in the test set not previously observed among the training units. When dealing with functional data, this requires learning the main patterns for the curves in the known classes, whilst being able to isolate signals that possess distinctive characteristics in the unlabeled set. In order to tackle this challenging problem, we propose a two-stage Bayesian semi-parametric novelty detector [2]. In the first stage, robust estimates are extracted from the training set via the Minimum Regularized Covariance Determinant (MRCD) estimator [1]. In the second stage, such information is employed to elicit informative priors within a Bayesian mixture of known groups plus a novelty term. To reflect the lack of knowledge on the latter component, we resort to a Dirichlet Process mixture model, thus overcoming the problematic a-priori specification of the expected number of novelties that may be present in the test set. The described methodology is applied to a spectroscopic dataset within a food authenticity study.

Keywords: bayesian mixture model, dirichlet process mixture model, functional data, minimum regularized covariance determinant

References

1. Denti, F., Cappelozzo, A., Greselin, F.,: A two-stage Bayesian semiparametric model for novelty detection with robust prior information. *Stat. Comput.* **31**, 42 (2021)
2. Boudt, K., Rousseeuw, P.J., Vanduffel, S., Verdonck, T.: The minimum regularized covariance determinant estimator. *Stat. Comput.* **30**, 113–128 (2020)

Francesco Denti
Università Cattolica del Sacro Cuore, Largo Agostino Gemelli, 1, 20123 Milano
e-mail: francesco.denti@unicatt.it

Andrea Cappelozzo
MOX - Laboratory for Modeling and Scientific Computing, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, e-mail: andrea.cappelozzo@polimi.it

Francesca Greselin
University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milano
e-mail: francesca.greselin@unimib.it



IFCS 2022

**17th Conference of the International Federation of Classification Societies
Classification and Data Science in the Digital Age**

Book of Abstracts

