IFCS 2022

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 V Invited Sessions

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

Francesco Denti, Andrea Cappozzo, 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., Cappozzo, 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 Cappozzo
MOX - Laboratory for Modeling and Scientific Computing, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, e-mail: andrea.cappozzo@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