EDITORIAL



Classification, sorting and clustering methods based on multiple criteria: recent trends

Salvatore Corrente¹ · Yves De Smet² · Michalis Doumpos³ · Salvatore Greco¹ · Constantin Zopounidis^{3,4}

Published online: 25 April 2023 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

A wide range of decision-making problems in various areas of engineering and management require the assignment of a set of decision options (alternatives) to categories/classes or groups. Depending on the characteristics of these categories/classes/groups and whether they are predefined or not, such problems are referred to as classification/sorting or clustering. Classification and sorting problems describe situations where the existing options are assumed to follow predefined patterns that correspond to a set of well-defined categories/ classes, which can be preferentially ordered (sorting) or not (classification). On the other hand, clustering problems involve cases where one is interested in describing the underlying latent structure of the decision alternatives by organizing them into homogeneous clusters, that are not defined a priori.

Classification/sorting and clustering is a fundamental topic in artificial intelligence (AI), with several machine learning (ML) algorithms now being available for supervised (classification) and unsupervised learning (clustering) tasks. During the past two decades, these problems have gained much interest in multiple criteria decision aiding (MCDA), which adopts a constructive decision aiding perspective in contrast to the statistical pattern recognition approach typically adopted in AI/ML. MCDA methodologies in this area provide new tools for decision and preference modelling, as well as innovative methodologies to develop classification/sorting and clustering models based on multiple decision criteria and modelling factors. Moreover, recently there has been a growing interest in exploring the interactions and synergies between the MCDA paradigm with AI/ML approaches, for preference learning and data-driven decision aiding.

Constantin Zopounidis kzopounidis@tuc.gr

- ² Unité CoDE-SMG, Université libre de Bruxelles, Avenue F.D. Roosevelt 50, Brussels 1050, Belgium
- ³ Technical University of Crete, School of Production Engineering and Management, Financial Engineering Laboratory, University Campus, Chania 73100, Greece
- ⁴ Audencia Business School, Nantes, France

¹ Department of Economics and Business, University of Catania, Corso Italia 55, Catania 95129, Italy

This context has been the motivation for the preparation of this special issue, which seeks to present the recent advances and trends in MCDA approaches for classification/sorting and clustering. Overall, 63 papers were submitted to the special issue and after a rigorous review process 20 papers were accepted for publication. The selected papers cover various methodological developments and applications in areas such as logistics and transport, urban planning, environmental assessment, energy efficiency analysis, and financial decision-making. Below we provide an outline of the special issue's contents.

The special issue starts with the paper by Ben Amor, Belaid, Benkraiem, Ramdani, and Guesmi. The paper presents a bibliographic analysis of multicriteria sorting and clustering methods, highlighting influential studies in these areas, identifying the status in the field, and pointing recent research developments and trends.

The second paper by Minoungou, Mousseau, Ouerdane, and Scotton, focuses on the Majority Rule Sorting (MR-Sort) method and the inference of the method's parameters from a set of decision examples. A mixed-integer optimization approach is presented that allows to learn a MR-Sort model from data. A unique aspect of the proposed approach is that it allows the consideration of cases with non-monotonic criteria.

In the third paper, Fernández, Figueira, and Navarro consider the development of ordinal classification methods based on comparing actions with limiting boundaries of ordered categories. The authors present new assignment procedures based on reflexive and asymmetric preference relations. The methods introduced in the paper are applicable under several conditions on boundaries of the categories and their structural properties are analyzed.

A similar problem is also examined in the next paper. More specifically, Raboun, Chojnacki, and Tsoukiàs present a new method for sorting (rating) problems, which aggregates "positive" and "negative" reasons supporting the assignments of the alternatives to specific categories. The proposed method is based on comparisons against the profiles characterizing the categories as well as comparisons among the alternatives.

The paper of Qin, Liang, Martinez, Ishizaka, and Pedrycz, introduces the ORESTE-SORT method for multicriteria sorting problems, which is based on the principles of outranking relations theory. The paper analyzes the properties of the method, including a procedure to derive the importance of the criteria based on the Besson rule, as well as a new rule for assigning the alternatives to the categories. The applicability of the new method is illustrated through a case study involving the competitiveness of ports in China.

The sixth paper by Zhang and Li, considers sorting problems in the context of group decision-making. The authors present two algorithms based on the TOPSIS method to reach consensus results. The algorithms facilitate the derivation of consensual boundary profiles as well as consensus results for the assignment of the alternatives to predefined categories. The applicability of the proposed algorithms is illustrated through an example of green building rating, as well as through simulation experiments.

In the next paper, Babashahi, Hansen, and Peeters examine the elicitation and aggregation of preference data collected through surveys, regarding the relative importance of the criteria. The authors present a method for evaluating the external validity of preference data obtained from non-random sampling based on measuring the cohesiveness of the data from the various exogenously defined groups in the sample. To this end, proper distance measures are presented combined with a bootstrapping method that allows for statistical testing. The proposed approaches are applied to a survey involving the evaluation of non-communicable diseases according to their overall burden to society. The paper by Cavallo and Ishizaka examines the scales used to express the decision makers' linguistic assessments in multicriteria methods based on pairwise comparisons. The authors conduct an experimental analysis to examine which type of scale better matches the decision maker's assessments. Eight scales are compared for problems of different sizes.

Ciardiello and Genovese present a comparative analysis between the TOPSIS multicriteria method and the simple additive weighting model, focusing on decision-making problems where the number of alternatives is much larger than the number of criteria. For the analysis, the authors consider different distance metrics for TOPSIS. Experimental results from the application of the two methods on simulated data, lead to insights on the similarities of the two approaches.

In the next paper, Xu, Chang, and Liu consider the development of data-driven decision models for classification problems, through an ensemble learning approach. The proposed methodology relies on a two-stage optimization model for assigning weights to a set of base classifiers, first by minimizing the ensemble error and then by maximizing the diversity of the ensemble. Moreover, to obtain interpretable results, a model to derive the relative importance of the criteria is also employed. The performance of the methodology is examined through a medical diagnosis application as well as through experiments on various benchmark datasets.

An ensemble approach for constructing classification models is also examined in the paper of Jha and Cucculelli. The authors propose novel multi-objective optimization strategies to specify the weights of a set of base models in a linear combination setting. The performance of the proposed methodology is tested on two real datasets involving credit risk and business innovation.

The paper by Norese, Rolando, and Curto introduces a knowledge-based multicriteria methodology that integrates logical and analytical tools to organize dispersed knowledge in complex problems when a decision process has not yet been activated. The proposed methodology first provides insights into the problem and then enables an interactive process to be applied in a participative context to explore possible actions. The methodology is illustrated through a pilot study involving the enhancement process of an urban system in Italy.

Pinto, Ferreira, Spahr, Sunderman, and Pereira apply a multicriteria methodology that combines cognitive mapping with the DEMATEL method to analyze blight and identify its causal factors. Cognitive mapping is employed to identify, select, prioritize, and group causes of blight based on a series of consultations with a panel of experts involving a case study in Portugal. In a second stage, the DEMATEL method is used to perform a quantitative analysis of blight's causes, facilitating prioritization and classification of these causes and the clarification of their interrelationships' dynamics.

In the next paper, Ren, Zhou, Makowski, Zhang, Yu, and Ma, examine the problem of prioritizing combinations of technologies for improving energy efficiency in the iron and steel production industry. The proposed approach is interactive, and it relies on a multicriteria programming model based on an achievement satisfaction function to determine the optimal technology adoption strategy. The proposed approach is implemented on data from the iron and steel industry in China.

Jose, Agarwal, Zhuang, and Swaminathan present a novel hierarchical fuzzy axiomatic design method to evaluate the performance of railway systems. The approach developed in the paper combines tangible and intangible criteria that are organized in a hierarchical manner. The methodology is applied to the evaluation of 16 zones in the Indian Railways.

Tlili, Khaled, Mousseau, and Ouerdane propose a multiobjective interactive approach based on a constrained non-compensatory sorting model for portfolio selection. The methodology operates on two evaluation levels, with the first focusing on screening the alternatives and constructing portfolios based on the decision maker's preferences, whereas the second level focuses on portfolio evaluation.

Roy, Shaw, and Ishizaka propose a fuzzy multicriteria approach for developing a credit rating system for small- and medium-sized enterprises. The methodology combines the fuzzy best-worst method (BWM) with a new fuzzy-TOPSIS sorting approach, into an expert-driven decision support system that considers financial and non-financial criteria.

Gaganis, Papadimitri, Pasiouras, and Tasiou present an empirical analysis on the use of social traits in predicting credit card delinquencies. The study follows a two-stage framework, first segmenting a market into homogeneous sub-populations at the regional level in terms of social traits, and then constructing delinquency prediction models for the identified segments. The results presented for a big dataset of over three million credit card holders in 12 UK regions, demonstrate that the proposed segmentation approach yields improved results compared to prediction in the overall population.

Hwang, Lee, and Fabozzi use deep clustering techniques to identify the characteristics of households' finances in a multidimensional context. The proposed approach combines a dimension reduction component and a clustering algorithm based on Gaussian mixture model. The methodology is applied to a dataset obtained from a survey on the financial soundness of Korean households and comparative results are presented against other clustering algorithms.

The special issue closes with the paper by du Jardin, who presents the development of a classification approach for bankruptcy prediction, based on convolutional neural networks. The author introduces a new topological transformation of financial data using a self-organizing map, which enables the derivation of a distance measure between the financial profile of a given company and the quantification of profiles belonging to a set of companies. The performance of the methodology is compared to other popular approaches for bankruptcy prediction using a dataset of French companies.

Closing this editorial, we express our sincere thanks to all authors whose contributions have been essential in completing this rich and high-quality special volume. We also acknowledge the support of all reviewers who devoted considerable time to provide critical evaluations, insightful comments, and constructive suggestions for the submitted papers. Without their help it would be impossible to achieve this volume's high standards.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.