EURO Journal on Computational Optimization 12 (2024) 100097

Contents lists available at ScienceDirect

EURO Journal on Computational Optimization

journal homepage: www.elsevier.com/locate/ejco

Modern optimization approaches to classification—Special issue editorial

1. Introduction

The last few years have witnessed an increasing interaction between the Optimization and Data Science communities, leading to innovative results in both fields. This special issue is related to the topics of the 17th International Conference of the International Federation of Classification Societies - IFCS 2022. It is a collection of original research papers reflecting the intersection of optimization and data science topics. The interplay between these areas, in fact, is essential in the current scenario where the development of complex, data-analytic systems plays a fundamental role. Different supervised and weakly supervised classification paradigms are considered, like optimal time shapelets for time series, logistic regression, Multiple Instance Learning, Graph Semi-Supervised learning. In all the papers, important aspects like interpretability and transparency are considered, proposing new algorithmic techniques and/or new models. The high quality of the submitted papers is shown by the Marguerite Frank Award for the best EJCO paper 2023 [1], won by the paper [2] where a new classification method based on a cloud of spheres is proposed.

We report more in details the contents of the seven papers published in the special issue.

In the paper [3] the authors try to solve the problem of feature subset selection in logistic regression using modern optimization techniques. The problem is formulated as a mixed-integer exponential cone program (MIEXP). Both nonlinear and discrete aspects of the underlying problem are considered within an exact optimization framework evaluated over a set of toy examples and benchmark datasets. The results show that the approach is quite successful in obtaining accurate and interpretable prediction models compared to other methods from the literature.

In the paper [4] the weakly supervised problem of binary Multiple Instance Learning is explored, where each sample is represented by a bag of instances. The class labels are provided only for each bag, whereas the instance labels are unknown: a bag is considered positive if at least one of its instances is positive and it is considered negative otherwise. In the paper a maximum-margin polyhedral separation surface is generated such that, for each positive bag, at least one of its instances is inside the polyhedron and all the instances of the negative bags are outside. The resulting optimization problem is a nonlinear, nonconvex and nonsmooth mixed-integer program, that is heuristically solved by a Block Coordinate Descent type method, based on repeatedly applying the DC (Difference of Convex) Algorithm.

In the paper [2] a binary classification model is proposed to distinguish a specific class. The model is based on a cloud of spheres that circumscribes the points of the class to be identified, and this condition is enforced trough the connectivity of a graph induced by the spheres. A quadratic problem with continuous and binary variables (MINLP) is proposed to build the model, having as objective the minimization of the number of spheres. This classification tool is effective when the structure of the class to be identified is highly non-linear and non-convex. Unlike neural networks, the classification model is transparent, with the structure perfectly identified. No kernel functions are used and it is not necessary to use meta-parameters unless it is intended also to maximize the separation margin as it is done in SVM. A heuristic is proposed that demonstrates nice results on a set of frequently tested real problems when compared to state-of-the-art algorithms.

In the paper [5] Graph Semi-Supervised learning is considered, where given a graph and a set of labeled nodes, the aim is to infer the labels to the remaining unlabeled nodes. An optimization-based formulation of the problem is proposed for undirected graphs, then extended to multilayer hypergraphs. The resulting problem is solved by different coordinate descent approaches. Experiments on synthetic and real-world datasets show the potential of using coordinate descent methods with suitable selection rules.

https://doi.org/10.1016/j.ejco.2024.100097

Available online 14 July 2024

The paper [6] develops an interpretable representation of a tree-ensemble model that can provide valuable insights into its behavior. First a hierarchical visualization tool is developed, based on a heatmap representation of the forest's feature use, considering the frequency of a feature and the level at which it is selected as an indicator of importance. Next, a mixed-integer linear programming (MILP) formulation is porposed for building a single optimal multivariate tree that accurately mimics the target model predictions. The goal is to provide an interpretable surrogate model based on oblique hyperplane splits, which uses only the most relevant features according to the defined forest's importance indicators. Computational experience shows that the proposed model is effective in yielding a shallow interpretable tree approximating the tree-ensemble decision function.

In the paper [7] a stochastic gradient method for solving the supervised classification problem is proposed. To control the variance of the objective's gradients, an automatic sample size selection along with a variable metric is used to precondition the stochastic gradient directions. Further, a non-monotone line search is defined to automatize step size selection. Convergence results are provided for both convex and non-convex objective functions. Extensive numerical experiments verify that the suggested approach performs on par with state-of-the-art methods for training both statistical models for binary classification and artificial neural networks for multi-class image classification.

The paper [8] deals with time series shapelets, which are a state-of-the-art data mining technique that is applied to time series supervised classification tasks. The main advantage of shapelets-based methods consists of their great interpretability. A novel mixed-integer Programming model to optimize shapelets discovery is proposed that is based on optimal binary decision trees. Computational results for a large class of datasets show that the approach achieves performance comparable with state-of-the-art shapelets-based classification methods.

References

[1] I. Bomze, The marguerite frank award for the best EJCO paper 2023, EURO J. Comput. Optim. 12 (2024) 100087, https://doi.org/10.1016/j.ejco.2024.100087. https://www.sciencedirect.com/science/article/pii/S2192440624000042

[2] T. Dias, P. Amaral, A classification method based on a cloud of spheres, EURO J. Comput. Optim. 11 (2023) 100077, https://doi.org/10.1016/j.ejco.2023.100077.https://www.sciencedirect.com/science/article/pii/S2192440623000217

[3] S. Asgharieh Ahari, B. Kocuk, A mixed-integer exponential cone programming formulation for feature subset selection in logistic regression, EURO J. Comput. Optim. 11 (2023) 100069, https://doi.org/10.1016/j.ejco.2023.100069.https://www.sciencedirect.com/science/article/pii/S2192440623000138

[4] A. Astorino, M. Avolio, A. Fuduli, Maximum-margin polyhedral separation for binary multiple instance learning, EURO J. Comput. Optim. 11 (2023) 100070, https://doi.org/10.1016/j.ejco.2023.100070.https://www.sciencedirect.com/science/article/pii/S219244062300014X

[5] S. Venturini, A. Cristofari, F. Rinaldi, F. Tudisco, Laplacian-based semi-supervised learning in multilayer hypergraphs by coordinate descent, EURO J. Comput. Optim. 11 (2023) 100079, https://doi.org/10.1016/j.ejco.2023.100079.https://www.sciencedirect.com/science/article/pii/S2192440623000230

[6] G. Di Teodoro, M. Monaci, L. Palagi, Unboxing tree ensembles for interpretability: a hierarchical visualization tool and a multivariate optimal re-built tree, EURO J. Comput. Optim. 12 (2024) 100084, https://doi.org/10.1016/j.ejco.2024.100084.https://www.sciencedirect.com/science/article/pii/S2192440624000017

[7] P. Cascarano, G. Franchini, E. Kobler, F. Porta, A. Sebastiani, A variable metric proximal stochastic gradient method: an application to classification problems, EURO J. Comput. Optim. 12 (2024) 100088, https://doi.org/10.1016/j.ejco.2024.100088.https://www.sciencedirect.com/science/article/pii/S2192440624000054

[8] L. Bonasera, S. Gualandi, Optimal shapelets tree for time series interpretable classification, EURO J. Comput. Optim. 12 (2024) 100091, https://doi.org/10.1016/j.ejco.2024.100091.https://www.sciencedirect.com/science/article/pii/S219244062400008X

António Pedro Duarte Silva[a,*], Laura Palagi[b], Veronica Piccialli[b]
[a] *Universidade Católica Portuguesa, Catolica Porto Business School, Portugal*
[b] *DIAG- Sapienza University of Rome, Italy*

[*] Corresponding author.
E-mail address: psilva@ucp.pt (A.P. Duarte Silva).