Summary of the work entitled

On optimal regression trees to detect critical intervals for multivariate functional data

State of the art

Extracting knowledge from data is a crucial task in Statistics and Machine Learning, and is at the core of many fields, such as Criminal Justice [34], Health Care [4] and Risk Management [2]. Mathematical Optimization plays an important role in building such models and interpreting their output, see, e.g., [17] for a recent survey.

Classification and regression trees [27] are state-of-the-art methods based on recursive partitioning. They show excellent learning performance, are conceptually simple, appealing in terms of interpretability because of their rule-based nature, computationally cheap and routines and packages to train them are available in popular languages such as Python and R.

The main goal of a classification and regression tree is to predict, as accurately as possible, the response variable using the predictor variables. On the top of this primary goal, other important characteristics may need to be considered, such as, e.g., cost-sensitivity constraints to protect risk groups [24]; fairness of the method avoiding the discrimination of groups that share sensitive features such as gender and race [1]; and explainability properties, e.g., sparsity of the tree model [6] and local interpretability of the model [28].

Since constructing optimal binary classification and regression trees is known to be an NP-complete problem [22], early research traditionally focused on the design of greedy heuristic procedures that require a low computational effort, e.g., CART [12]. These models cannot include the desirable global structural properties mentioned above due to their greedy nature.

In recent times, and because of the availability of more powerful hardware and the dramatic advances in optimization solvers, there has been an increased interest by the Mathematical Optimization community to develop novel approaches to build optimal decision trees. These new paradigms include Mixed-Integer Linear Optimization (MILO) [7, 16, 20, 31], Continuous Optimization [9, 10, 11], Constraint Programming [30], SAT [29] and Dynamic Programming [15] approaches, as reviewed in [13]. Contrary to greedy approaches, formulating the design of the tree model as an optimization problem allows one to easily include, either as hard or soft constraints, desirable global structural properties.

Optimal decision trees have mainly focused on the analysis of multivariate data. Nevertheless, other kinds of complex data, such as functional data, are also of interest. Functional Data Analysis [14, 18, 33] is the field of Statistics that extends the classic multivariate analysis to handle observations of functional nature, and has applications in areas such as Biomedicine [26], Meteorology [23] and Econometrics [25].

In principle, optimal decision trees, as any other standard multivariate approach, could address the analysis of functional data after discretizing the functions and converting them to vectors. Yet, in general, the direct use of such techniques for functional data may have dramatic consequences, for instance, the curse of dimensionality since lots of points are needed to summarize the information in a curve [32]. Furthermore, the intrinsic characteristics of functional data may not be fully exploited: multivariate approaches ignore the ordering and the spacing of a set of data values [19], and the strong relationship between measurements in two consecutive instants is not taken into account, potentially yielding multicollinearity problems [21]. For these reasons, it is preferable to develop models that take advantage of the functional nature directly [3].

Summary

In this work, we propose a Continuous Optimization formulation, based on previous research of the authors [9, 10, 11, 13], to build regression trees that handle multivariate functional data. It may happen that simply using a finite number of instants [5] and/or intervals [8] in the domain of the functional predictor variables is sufficient to produce accurate predictions. Our model controls explicitly the number of intervals that are critical for prediction, as well as their length, and assigns to each of them a single coefficient per functional variable that is constant for the whole interval. This has, in turn, the advantage of being more interpretable and saves both monitoring and storage costs. In the case of degenerate intervals with length equal to zero, critical instants would be detected instead.

Contribution to the literature

This work contributes to the literature in several directions. First, we address multivariate functional data through optimal regression trees that are competitive in terms of prediction accuracy in real-world data sets. Second, our proposal is scalable with respect to the size of the training sample, since there are no decision variables directly relating to the individuals, as opposed to MILO approaches to build optimal trees. Third, the sparsity of the model is achieved through the inclusion of regularization terms that control the number of non-zero coefficients, the number of predictor variables and, in the case of functional ones, the proportion of the domain used for prediction. Fourth, our approach is flexible allowing the incorporation of fairness and cost-sensitivity contraints. Finally, thanks to the smoothness in the prediction function of our approach, a full sensitivity analysis can be performed in order to study the impact that continuous predictor variables have on each individual prediction, thus enhancing the local interpretability of tree models.

References

[1] S. Aghaei, M. Azizi, and P. Vayanos. Learning optimal and fair decision trees for non-discriminative decision-making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1418–1426, 2019.

[2] B. Baesens, R. Setiono, C. Mues, and J. Vanthienen. Using neural network rule extraction and decision tables for credit-risk evaluation. 49(3):312–329, 2003.

[3] S. Balakrishnan and D. Madigan. Decision trees for functional variables. In *Sixth International Conference on Data Mining (ICDM'06)*, pages 798–802, 2006.

[4] S. Benítez-Peña, E. Carrizosa, V. Guerrero, M. D. Jiménez-Gamero, B. Martín-Barragán, C. Molero-Río, P. Ramírez-Cobo, D. R. Morales, and M. R. Sillero-Denamiel. On sparse ensemble methods: An application to short-term predictions of the evolution of COVID-19. *European Journal of Operational Research*, 295(2):648–663, 2021.

[5] J. R. Berrendero, B. Bueno-Larraz, and A. Cuevas. An RKHS model for variable selection in functional linear regression. *Journal of Multivariate Analysis*, 170:25–45, 2019.

[6] D. Bertsimas and V. Digalakis. The backbone method for ultra-high dimensional sparse machine learning. *Machine Learning*, pages 1–52, 2022.

[7] D. Bertsimas and J. Dunn. Optimal classification trees. *Machine Learning*, 106(7):1039–1082, 2017.

[8] R. Blanquero, E. Carrizosa, A. Jiménez-Cordero, and B. Martín-Barragán. Selection of time instants and intervals with support vector regression for multivariate functional data. *Computers & Operations Research*, 123:105050, 2020.

[9] R. Blanquero, E. Carrizosa, C. Molero-Río, and D. Romero Morales. Sparsity in optimal randomized classification trees. *European Journal of Operational Research*, 284(1):255 – 272, 2020.

[10] R. Blanquero, E. Carrizosa, C. Molero-Río, and D. Romero Morales. Optimal randomized classification trees. *Computers & Operations Research*, 132:105281, 2021.

[11] R. Blanquero, E. Carrizosa, C. Molero-Río, and D. Romero Morales. On sparse optimal regression trees. *European Journal of Operational Research*, 299(3):1045–1054, 2022.

[12] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC Press, 1984.

[13] E. Carrizosa, C. Molero-Río, and D. Romero Morales. Mathematical optimization in classification and regression trees. *TOP*, 29(1):5–33, 2021.

[14] A. Cuevas. A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147:1–23, 2014.

[15] E. Demirović, A. Lukina, E. Hebrard, J. Chan, J. Bailey, C. Leckie, K. Ramamohanarao, and P. J. Stuckey. MurTree: Optimal Classification Trees via Dynamic Programming and Search. *Journal of Machine Learning Research*, 23(26):1–47, 2022.

[16] M. Firat, G. Crognier, A. Gabor, C. Hurkens, and Y. Zhang. Column generation based math-heuristic for classification trees. *Computers & Operations Research*, 116:104866, 2019.

[17] C. Gambella, B. Ghaddar, and J. Naoum-Sawaya. Optimization problems for machine learning: A survey. *European Journal of Operational Research*, 290(3):807–828, 2021.

[18] A. Goia and P. Vieu. An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, 146:1–6, 2016.

[19] C. K. Griswold, R. Gomulkiewicz, and N. Heckman. Hypothesis testing in comparative and experimental studies of function-valued traits. *Evolution*, 62(5):1229–1242, 2008.

[20] O. Günlük, J. Kalagnanam, M. Li, M. Menickelly, and K. Scheinberg. Optimal Decision Trees for Categorical Data via Integer Programming. *Journal of Global Optimization*, 81:233–260, 2021.

[21] T. Hastie, A. Buja, and R. Tibshirani. Penalized discriminant analysis. *The Annals of Statistics*, 23(1):73–102, 1995.

[22] L. Hyafil and R. Rivest. Constructing optimal binary decision trees is NP-complete. *Information Processing Letters*, 5(1):15–17, 1976.

[23] G. M. James, J. Wang, and J. Zhu. Functional linear regression that's interpretable. *The Annals of Statistics*, 37(5A):2083–2108, 2009.

[24] H.-P. Kao and K. Tang. Cost-sensitive decision tree induction with label-dependent late constraints. *INFORMS Journal on Computing*, 26(2):238–252, 2014.

[25] A. Laukaitis and A. Račkauskas. Functional data analysis for clients segmentation tasks. *European Journal of Operational Research*, 163(1):210–216, 2005.

[26] X. Leng and H.-G. Müller. Classification using functional data analysis for temporal gene expression data. *Bioinformatics*, 22(1):68–76, 2005.

[27] W.-Y. Loh. Fifty years of classification and regression trees. *International Statistical Review*, 82(3):329–348, 2014.

[28] S. Lundberg, G. Erion, H. Chen, A. DeGrave, J. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1):2522–5839, 2020.

[29] N. Narodytska, A. Ignatiev, F. Pereira, and J. Marques-Silva. Learning Optimal Decision Trees with SAT. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, pages 1362–1368, 2018.

[30] H. Verhaeghe, S. Nijssen, G. Pesant, C.-G. Quimper, and P. Schaus. Learning optimal decision trees using constraint programming. In *The 25th International Conference on Principles and Practice of Constraint Programming (CP2019)*, 2019.

[31] S. Verwer, Y. Zhang, and Q. Ye. Learning optimal classification trees using a binary linear program formulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1625–1632, 2019.

[32] P. Vieu. On dimension reduction models for functional data. *Statistics & Probability Letters*, 136:134–138, 2018.

[33] J.-L. Wang, J.-M. Chiou, and H.-G. Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

[34] J. Zeng, B. Ustun, and C. Rudin. Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society: Series A*, 180(3):689–722, 2017.