Firat, Murat; Crognier, Guillaume; Gabor, Adriana F.; Hurkens, C. A. J.; Zhang, Yingqian
Column generation based heuristic for learning classification trees. (English) Zbl 1458.68201

Summary: This paper explores the use of Column Generation (CG) techniques in constructing univariate binary decision trees for classification tasks. We propose a novel Integer Linear Programming (ILP) formulation, based on root-to-leaf paths in decision trees. The model is solved via a Column Generation based heuristic. To speed up the heuristic, we use a restricted instance data by considering a subset of decision splits, sampled from the solutions of the well-known CART algorithm. Extensive numerical experiments show that our approach is competitive with the state-of-the-art ILP-based algorithms. In particular, the proposed approach is capable of handling big data sets with tens of thousands of data rows. Moreover, for large data sets, it finds solutions competitive to CART.

MSC:
68T20 Problem solving in the context of artificial intelligence (heuristics, search strategies, etc.)
62H30 Classification and discrimination; cluster analysis (statistical aspects)
68T05 Learning and adaptive systems in artificial intelligence
68T09 Computational aspects of data analysis and big data
90C10 Integer programming

Keywords:
machine learning; decision trees; column generation; classification; CART; integer linear programming

Software:
UCI-ml; CPLEX

Full Text: DOI arXiv

References:
[7] 1-32 · Zbl 1165.90578


[18] Lehigh University

[19] Volume 2


[27] CPAIOR 2017, Padua, Italy, June 5-8, 2017, Proceedings - Zbl 06756578


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.