Learning Optimal Decision Trees using Constraint Programming

Hélène Verhaeghe[1], Siegfried Nijssen[1], Gilles Pesant[2], Claude-Guy Quimper[3], and Pierre Schaus[1]

[1]UCLouvain, ICTEAM, Place Sainte Barbe 2, 1348 Louvain-la-Neuve, Belgium, {firstname.lastname}@uclouvain.be
[2]Polytechnique Montréal, Montréal, Canada, gilles.pesant@polymtl.ca
[3]Université Laval, Québec, Canada, claude-guy.quimper@ift.ulaval.ca

Abstract. Decision trees are among the most popular classification models in machine learning. Using greedy algorithms to learn them can pose several disadvantages: it is difficult to limit the size of the decision trees while maintaining a good classification accuracy, and it is hard to impose additional constraints on the models that are learned. For these reasons, there has been a recent interest in exact and flexible algorithms for learning decision trees. This paper is a summary of our paper "Learning Optimal Decision Trees using Constraint Programming" accepted in CP2019 [4]. In our paper, we introduce a new approach to learn decision trees using constraint programming.

Decision trees are popular classification models in machine learning. Benefits of decision trees include that they are relatively easy to interpret and that they provide good classification performance on many datasets.

Several methods have been proposed in the literature for learning decision trees. The greedy methods are the most popular ones. These methods recursively partition a dataset into two subsets based on a greedily selected attribute until some stopping criterion is reached (such as a minimum number of examples in the leaf, or a unique class label in these examples). While in practice these methods obtain a good prediction accuracy for many types of data, unfortunately, they provide little guarantees. As a result, the trees learned using these methods may be unnecessarily complex, may be less accurate than possible, and it is hard to impose additional constraints on the trees.

To address these weaknesses, researchers have studied the inference of *optimal* decision trees under constraints [2]. These approaches ensure that under well-defined constraints and optimization criteria, an optimal tree is found.

Our paper proposes a new, more scalable approach based on Constraint Programming (CP) for learning decision trees with a fixed maximum depth minimizing the classification error. Our approach combines these key ideas: the use of branch-and-bound in a CP solver, the use of the CoverSize global constraint

[3], the use of an AND/OR search tree [1] and the use of caching as introduced in DL8 [2]. This allows our constraint programming approach to deal in a much more efficient way with the decompositions in the learning problem. We will show that the combination of these different ideas leads to a model that is more efficient than other approaches proposed in the literature.

The main decision variables required to model our decision tree are the decisions taken at each node. Another set of variables is used to count the transactions matching at each node of the tree. The integrity of the tree is ensured by the use of *AllDifferent* constraints, CoverSize constraints and simple arithmetic constraints linking the various variables. Other redundant constraints are added to help speed up the resolution.

We compared our methods to DL8 [2] and BinOCT [5], two methods addressing the same problem. Our method outperforms these two others on most of the instances. It could find and prove optimality on roughly 75% of the instances within the time limit. The best solution found was reached by our method in almost all cases. However, DL8 performs better on small instances. The large difference between BinOCT and our method can be explained by the benefits of the AND/OR search that is not used by BinOCT. The gap with DL8 can be partially explained by the cost pruning. It can potentially also be explained by the itemset mining algorithms used: DL8 lacks the optimizations found in the CoverSize constraint. Our experiments also evaluate the effects of some of the techniques used to solve the problem, such as the use of a cache.

In summary, our paper presents a new approach for efficiently creating an optimal decision tree of limited depth. This approach based on CP combines the CoverSize global constraint, the concept of AND/OR tree and caching. On most of the benchmarks, it gives the best solution within the allocated time and is the fastest to prove optimality. We believe our approach can be extended in many different ways (multiclass, continuous features through binarization,...).

References

1. Dechter, R., Mateescu, R.: And/or search spaces for graphical models. Artificial intelligence **171**(2-3), 73–106 (2007)
2. Nijssen, S., Fromont, E.: Mining optimal decision trees from itemset lattices. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 530–539. ACM (2007)
3. Schaus, P., Aoga, J.O., Guns, T.: Coversize: a global constraint for frequency-based itemset mining. In: International Conference on Principles and Practice of Constraint Programming. pp. 529–546. Springer (2017)
4. Verhaeghe, H., Nijssen, S., Pesant, G., Quimper, C.G., Schaus, P.: Learning optimal decision trees using constraint programming. In: Principles and Practice of Constraint Programming - CP 2019, 25th International Conference, Stamford, USA, September 30 - October 4, 2019 (2019)
5. Verwer, S., Zhang, Y.: Learning optimal classification trees using a binary linear program formulation. In: 33rd AAAI Conference on Artificial Intelligence (2019)