EasyChair Preprint

№ 2572

# Reinforcement Learning for Variable Selection in a Branch and Bound Algorithm

Marc Etheve, Zacharie Alès, Côme Bissuel, Olivier Juan and Safia Kedad-Sidhoum



February 5, 2020

# Reinforcement Learning for Variable Selection in a Branch and Bound Algorithm

Marc Etheve[1,3] Zacharie Alès[2,3] Côme Bissuel[1] Olivier Juan[1]
Safia Kedad-Sidhoum[3]

[1] EDF R&D, France
{marc.etheve, come.bissuel, olivier.juan}@edf.fr
[2] ENSTA Paris, Institut Polytechnique de Paris, France
zacharie.ales@ensta-paris.fr
[3] CNAM Paris, CEDRIC, France
safia.kedad_sidhoum@cnam.fr

**Mots-clés** : *Reinforcement Learning, Mixed Integer Linear Programming, Neural Network, Branch and Bound, Branching Strategy*

## 1 Introduction

Mixed Integer Linear Programming (MILP) is an active field of research due to its tremendous usefulness in real-world applications. The most common method designed to solve MILP problems is the Branch and Bound (B&B) algorithm (see [1] for an exhaustive introduction). B&B is a general purpose procedure dedicated to solve any MILP instance, based on a divide and conquer strategy and driven by generic heuristics and bounding procedures.

Recently, a lot of attention has been paid to the interactions between MILP and Machine Learning. As pointed out in [2], learning methods may compensate for the lack of mathematical understanding of the B&B method and its variants ([3, 4]). The plethora of different approaches in this young field of research gives evidence of the variety of ways in which learning can be leveraged ([2]). Whether it is by Imitation Learning or by Reinforcement Learning (RL), these solutions are often limited by their scope : they seek to take decisions according to a local criterion. In the present work, we propose a novel approach based on Reinforcement Learning aiming at optimising a global criterion at the scale of the whole B&B tree. We learn a branching strategy from scratch, independent of any heuristic and guaranteeing optimality.

Our contributions are three-fold. First, we present the RL task of defining an agent aiming at minimising a global criterion. With that objective, we demonstrate that, under certain assumptions, a specific kind of value functions enforces the optimality of such criterion. Last, we propose to adapt known generic learning methods and Neural Network architectures to the Branch and Bound setting. We illustrate our proposed method on industrial problems.

## 2 General setting

We consider the case where MILP instances of a given problem are stochastically generated. In this setting, our aim is to learn a branching strategy designed for performing well in average on this problem and according to a specified metric. Let us call $\mathcal{D}$ the instances distribution, $\Pi_\theta$ the generator of branching decisions parametrised by $\theta$ and $\mu$ the considered metric. Our objective is then to find $\theta^*$ such that

$$\theta^* \in \arg\min_{\theta \in \Theta} \mathbb{E}_{p \sim \mathcal{D}} \left[ \mu \left( \Pi_\theta \left( p \right) \right) \right]. \tag{1}$$

## 3 Fitting for Minimising the SubTree Size (FMSTS)

A natural way to enforce objective (1) is through Reinforcement Learning (see [5] for an introduction), for instance by considering the branching strategy as a Deterministic Markov Decision Process. Using Proposition 1, we propose MFSTS as an Approximate Q-learning algorithm using the subtree size under a B&B node as a value function.
This value function has two great advantages. First, it is directly observable once the whole tree has been expanded. More importantly, when using DFS, it is a local criterion optimal with respect to the tree size used as the global metric $\mu$.

**Proposition 1** *When using Depth First Search (DFS) as node selection strategy, minimising the whole B&B tree size is achieved when any subtree is of minimal size.*

## 4 Adaptations to the Branch and Bound setting

To help us enforcing (1), we adapt three traditional components in modern Deep Reinforcement Learning techniques.
First, we modify the loss used in Stochastic Gradient Descent by weighting the Bellman's Residual proportionally to the inverse of the whole tree size. This allows us to compensate for the natural bias of learning towards difficult instances. Second, we adapt Prioritized Experienced Replay [6] to the exponential structure of binary trees. Last, we propose a new Neural Network architecture inspired from the Dueling architecture [7].

## 5 Experiments and Perspectives

We test our method on industrial problems provided by EDF and compare FMSTS to CPLEX. To ensure a fair comparison between branching strategies, we prevent CPLEX from using its presolve and cuts. We exhibit promising results on small-size problems, sometimes better than CPLEX.
The learning task becomes computationally heavy as the problems get more and more complicated. The issue of making the method scalable to bigger problems will be tackled in a future work. However, to our knowledge, it is the first time a pure Reinforcement Learning algorithm has successfully been used to perform variable selection in a B&B scheme.

## Références


[1] L. A. Wolsey, *Integer programming*. Wiley, 1998.

[2] Y. Bengio, A. Lodi, and A. Prouvost, "Machine learning for combinatorial optimization : a methodological tour d'horizon," *arXiv preprint arXiv :1811.06128*, 2018.

[3] C. Barnhart, E. L. Johnson, G. L. Nemhauser, M. W. Savelsbergh, and P. H. Vance, "Branch-and-price : Column generation for solving huge integer programs," *Operations research*, vol. 46, no. 3, pp. 316–329, 1998.

[4] J. E. Mitchell, "Branch-and-cut algorithms for combinatorial optimization problems," *Handbook of applied optimization*, vol. 1, pp. 65–77, 2002.

[5] R. S. Sutton and A. G. Barto, *Reinforcement learning : An introduction*. MIT press, 2018.

[6] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," *arXiv preprint arXiv :1511.05952*, 2015.

[7] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, "Dueling network architectures for deep reinforcement learning," *arXiv preprint arXiv :1511.06581*, 2015.