Mach Learn (2013) 92:403–429
DOI 10.1007/s10994-013-5369-0

Hypervolume indicator and dominance reward based multi-objective Monte-Carlo Tree Search

Weijia Wang · Michèle Sebag

Received: 21 January 2013 / Accepted: 24 April 2013 / Published online: 17 May 2013

Abstract Concerned with multi-objective reinforcement learning (MORL), this paper presents MOMCTS, an extension of Monte-Carlo Tree Search to multi-objective sequential decision making, embedding two decision rules respectively based on the hypervolume indicator and the Pareto dominance reward. The MOMCTS approaches are firstly compared with the MORL state of the art on two artificial problems, the two-objective Deep Sea Treasure problem and the three-objective Resource Gathering problem. The scalability of MOMCTS is also examined in the context of the NP-hard grid scheduling problem, showing that the MOMCTS performance matches the (non-RL based) state of the art albeit with a higher computational cost.

Keywords Reinforcement learning · Monte-Carlo Tree Search · Multi-objective optimization · Sequential decision making

1 Introduction

Reinforcement learning (RL) (Sutton and Barto 1998; Szepesvári 2010) addresses sequential decision making in the Markov decision process framework. RL algorithms provide guarantees of finding the optimal policies in the sense of the expected cumulative reward, relying on the thorough exploration of the state and action spaces. The price to pay for these optimality guarantees is the limited scalability of mainstream RL algorithms w.r.t. the size of the state and action spaces.

Recently, Monte-Carlo Tree Search (MCTS), including the famed Upper Confidence Tree algorithm (Kocsis and Szepesvári 2006) and its variants, has been intensively investigated to handle sequential decision problems. MCTS, notably illustrated in the domain of Computer-Go (Gelly and Silver 2007), has been shown to efficiently handle medium-size state and action search spaces through a careful balance between the exploration of the

Editors: Zhi-Hua Zhou, Wee Sun Lee, Steven Hoi, Wray Bunfine, and Hiroshi Motoda.

W. Wang (✉) · M. Sebag
LRI, CNRS UMR 8623 & INRIA-Saclay, Université Paris-Sud, 91405 Orsay Cedex, France
e-mail: wjch111@gmail.com