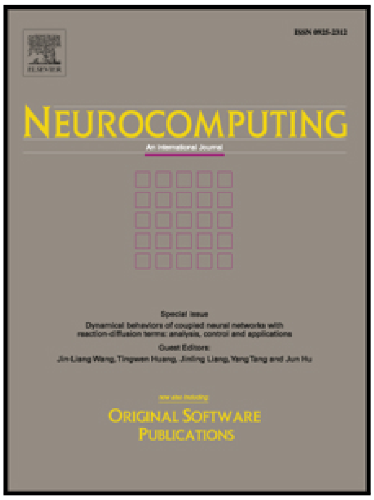
Accepted Manuscript

Steering Approaches to Pareto-Optimal Multiobjective Reinforcement Learning

Peter Vamplew, Rustam Issabekov, Richard Dazeley,
Cameron Foale, Adam Berry, Tim Moore, Douglas Creighton



PII: S0925-2312(17)31101-3
DOI: 10.1016/j.neucom.2016.08.152
Reference: NEUCOM 18602

To appear in: *Neurocomputing*

Received date: 3 March 2016
Revised date: 11 July 2016
Accepted date: 25 August 2016

Please cite this article as: Peter Vamplew, Rustam Issabekov, Richard Dazeley, Cameron Foale, Adam Berry, Tim Moore, Douglas Creighton, Steering Approaches to Pareto-Optimal Multiobjective Reinforcement Learning, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2016.08.152

Steering Approaches to Pareto-Optimal Multiobjective Reinforcement Learning

Peter Vamplew[a], Rustam Issabekov[a], Richard Dazeley[a], Cameron Foale[a], Adam Berry[b], Tim Moore[b], Douglas Creighton[c]

[a] *Federation Learning Agents Group, School of Engineering and Information Technology, Federation University Australia, Ballarat, Victoria, Australia*
[b] *Energy Technology Division, CSIRO, Mayfield West, NSW, Australia*
[c] *Centre for Intelligent Systems Research, Deakin University, Waurn Ponds, Victoria, Australia*

Abstract

For reinforcement learning tasks with multiple objectives, it may be advantageous to learn stochastic or non-stationary policies. This paper investigates two novel algorithms for learning non-stationary policies which produce Pareto-optimal behaviour (w-steering and Q-steering), by extending prior work based on the concept of geometric steering. Empirical results demonstrate that both new algorithms offer substantial performance improvements over stationary deterministic policies, while Q-steering significantly outperforms w-steering when the agent has no information about recurrent states within the environment. It is further demonstrated that Q-steering can be used interactively by providing a human decision-maker with a visualisation of the Pareto front and allowing them to adjust the agent's target point during learning. To demonstrate broader applicability, the use of Q-steering in combination with function approximation is also illustrated on a task involving control of local battery storage for a residential solar power system.

Keywords: multiobjective reinforcement learning, non-stationary policies, geometric steering, interactive reinforcement learning, Pareto optimality

1. Introduction

Reinforcement learning (RL) methods learn the optimal behaviour for an agent on the basis of a reward signal received from the agent's environment. While most RL research assumes the agent has only a single objective