

# 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](https://doi.org/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](https://doi.org/10.1016/j.neucom.2016.08.152)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Steering Approaches to Pareto-Optimal Multiobjective Reinforcement Learning

Peter Vamplew<sup>a</sup>, Rustam Issabekov<sup>a</sup>, Richard Dazeley<sup>a</sup>, Cameron Foale<sup>a</sup>,  
Adam Berry<sup>b</sup>, Tim Moore<sup>b</sup>, Douglas Creighton<sup>c</sup>

<sup>a</sup>*Federation Learning Agents Group, School of Engineering and Information Technology,  
Federation University Australia, Ballarat, Victoria, Australia*

<sup>b</sup>*Energy Technology Division, CSIRO, Mayfield West, NSW, Australia*

<sup>c</sup>*Centre for Intelligent Systems Research, Deakin University, Waurin 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

Download English Version:

<https://daneshyari.com/en/article/4947080>

Download Persian Version:

<https://daneshyari.com/article/4947080>

[Daneshyari.com](https://daneshyari.com)