



Pruning Dominated Policies in Multiobjective Pareto Q-learning*

***Note:** The full contents of this paper have been published in the volume *Lecture Notes in Artificial Intelligence 11160* (LNAI 11160)

Lawrence Mandow, José-Luis Pérez-de-la-Cruz
Departamento de Lenguajes y Ciencias de la Computación
Universidad de Málaga
Málaga, Spain
{lawrence, perez}@lcc.uma.es

Abstract—The solution for a Multi-Objective Reinforcement Learning problem is a set of Pareto optimal policies. MPQ-learning is a recent algorithm that approximates the whole set of all Pareto-optimal deterministic policies by directly generalizing Q-learning to the multiobjective setting. In this paper we present a modification of MPQ-learning that avoids useless cyclical policies and thus improves the number of training steps required for convergence.

Index Terms—