Scalarized Multi-Objective Reinforcement Learning:

Novel Design Techniques

Kristof Van Moffaert [a] Madalina M. Drugan [a] Ann Nowé [a]

[a] *Computational Modeling Lab, VUB, Pleinlaan 2, B-1050, Brussels*

Abstract

In multi-objective problems, it is key to find compromising solutions that balance different objectives. The linear scalarization function is often utilized to translate the multi-objective nature of a problem into a standard, single-objective problem. Generally, it is noted that such as linear combination can only find solutions in convex areas of the Pareto front, therefore making the method inapplicable in situations where the shape of the front is not known beforehand. We propose a non-linear scalarization function, called the Chebyshev scalarization function in multi-objective reinforcement learning. We show that the Chebyshev scalarization method overcomes the flaws of the linear scalarization function and is able to discover all Pareto optimal solutions in non-convex environments.

1 Introduction

Formally, multi-objective reinforcement learning (MORL) is the process of simultaneously optimizing multiple objectives which can be complementary, conflicting as well as independent. So deciding a priori on the importance of the different criteria might be difficult. The goal of MORL is to search the policy space and eventually find policies that simultaneously optimize one or more objectives.

A popular approach consists of transforming the multi-objective problem into a single-objective problem by employing *scalarization* functions. These functions provide a single score indicating the quality over a combination of objectives, which allows a simple and total ordering. In many cases, a linear combination of the objectives is utilized, but as noted in [1], this mechanism only allows Pareto optimal solutions to be found amongst convex areas of the Pareto front.

2 Scalarization functions

The linear scalarization function. In single-objective learning, the agent's table is used to store the expected reward for the combination of state s and action a, i.e. $\hat{Q}(s, a)$. In a multi-objective setting, the Q-table is extended to incorporate objectives, i.e. $\hat{Q}(s, a, o)$. Thus, the expected rewards for each state, action and objective can be stored, retrieved and updated separately.

An important aspect of multi-objective optimization consists of how the actions are selected, based on different objectives. A scalarization function transforms a multi-objective problem into a single objective problem by performing a function over the objectives to obtain a combined score for an action a for different objectives o. This single score can then be used to evaluate the particular action a. Given these scores, one can utilize the standard action selection strategies of single-objective reinforcement learning, such as ϵ-greedy and Boltzmann, to decide which action to select. Most scalarization functions imply that an objective

o is associated with a weighted coefficient, which allows the user some control over the nature of the policy found by the system, by placing greater or lesser emphasis on each of the objectives. In a multi-objective environment, this trade-off is parametrized by $w_o \in [0, 1]$ for objective o and $\sum_{o=1}^{m} w_o = 1$. The most common function is the linear scalarization function because of its simplicity and straightforwardness. More precisely, a weighted-sum is performed over each $\hat{Q}(s, a, o)$ with $o = 1 \dots m$ and their corresponding weights to obtain the score of x, i.e. As a result of applying the scalarization, scalarized Q-values or SQ-values are obtained: $SQ(s, a) = \sum_{o=1}^{m} w_o \cdot \hat{Q}(s, a, o)$.

The Chebyshev scalarization function. Our novel alternative as a mechanism to evaluate actions with multiple objectives consists of using L_p metrics. In detail, L_p metrics measure the distance between a point in the multi-objective space and a utopian point z^*. In our setting, we measure this distance to the value of the objective functions f for each objective o of the multi-objective solution x, i.e. $\min_{x \in \mathbb{R}^n} L_p(x) = \left(\sum_{o=1}^{m} w_o |f_o(x) - z_o^*|^p \right)^{1/p}$, where $1 \leq p \leq \infty$. In the case of p = ∞, the metric is also called the weighted L_∞ or the Chebyshev metric and is of the form: $\min_{x \in \mathbb{R}^n} L_\infty(x) = \max_{o=1 \dots m} w_o |f_o(x) - z_o^*|$ In terms of action selection mechanism, the objective function values f are replaced by $\hat{Q}(s, a, o)$-values to obtain the scalarized Q-value or SQ-value, for state s and action a: $SQ(s, a) = \max_{o=1 \dots m} w_o \cdot |\hat{Q}(s, a, o) - z_o^*|$ The reference point z^* is a parameter that is being constantly adjusted during the learning process by recording the best value so far for each objective o, plus a small constant τ, i.e. $z_o^* = f_o^{best}(x) + \tau$.

3 Experiments

We evaluated the two scalarization functions on two benchmark environments, called the Deep Sea Treasure world and the multi-objective Mountain Car world with two and three objectives respectively [1]. The policies found are evaluated by the hypervolume metric which measures the distance to the set of Pareto optimal solutions and the spread of the policies. The hypervolume measures in Fig. 1 teach us that the

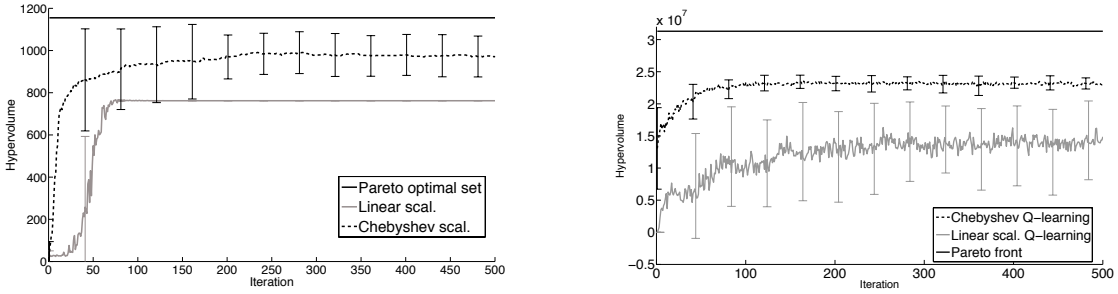

Figure 1: The learning curve of the Q-learners using the linear and Chebyshev scalarization functions as their action evaluation methods on the DST world. The Pareto optimal set is depicted in black.

non-linear Chebyshev function is able to attain a greater part of set of Pareto optimal solutions, while the linear scalarization function was limited to only obtaining a restricted set of policies. Other experiments also showed us that the Chebyshev function attained a much more diverse set of policies than the linear function in non-convex environments [2].

References

[1] Peter Vamplew, Richard Dazeley, Adam Berry, Rustam Issabekov, and Evan Dekker. Empirical evaluation methods for multiobjective reinforcement learning algorithms. *Machine Learning*, 84(1-2):51–80, 2010.

[2] Kristof. Van Moffaert, Madalina M. Drugan, and A. Nowé. Scalarized Multi-Objective Reinforcement Learning: Novel Design Techniques. In *2013 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*. IEEE, 2013.