Geophysical Research Abstracts
Vol. 13, EGU2011-876, 2011
EGU General Assembly 2011

Reinforcement Learning for Multi-Objective Water System Management

Francesca Pianosi, Andrea Castelletti, Stefano Galelli, Marcello Restelli, and Rodolfo Soncini-Sessa
Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy
({pianosi,castelle,galelli,restelli,soncini}@elet.polimi.it)

Despite the great progress made in the last decades, optimal management of water systems still remains a very active research area. Water resources management problems are sequential decision processes, meaning that decisions taken at a certain time not only produce an immediate reward, but also affect the next system state and thus the future space of feasible decisions. Finding the optimal management is especially difficult when the water system input are highly uncertain and a closed-loop (feedback) approach must be adopted, which means searching for an operating rule that can provide the optimal decision for any possible state encountered by the system.
Stochastic Dynamic Programming (SDP) is one of the most widely used methods for designing optimal operating rules. It is based on the iterative approximation of the optimal value function by balancing immediate reward and long-term ones, as measured by the optimal value function from previous iteration. However, SDP suffers from a dual curse that prevents its practical application to even reasonably complex water systems: the so-called curse of dimensionality (computational complexity grows exponentially with state, decision and random input dimension); and the curse of modelling (an explicit model of each water system component is required to anticipate the effects of the system transition). Computational burden of SDP is accrued when dealing with multi-objective (MO) problems, because the (Pareto) optimal value function must be repeatedly computed for different objective tradeoffs (weighting method).
In this paper we propose a novel Reinforcement Learning (RL) approach, called Multi-Objective Fitted Q-Iteration (MOFQI), which combines RL concepts of learning from experience and continuous approximation of the value function, to simultaneously cope with the twin curses of modelling and dimensionality, and especially suited for efficiently handling MO problems.
The advantages of MOFQI are that: (i) the learning process can be performed off-line without directly experimenting on the real system, which would cause unsustainable environmental and economic costs; (ii) the learning process is based on a dataset of system transition samples and associated rewards, which can be obtained by model simulation as well as directly from historical time series; (iii) through efficient sampling of the state-control space, it gets the same level of accuracy as SDP while using a definitely coarser discretization grid and thus a much reduced computing time; (iv) by enlarging the continuous approximation of the value function to the weight space, it can derive the entire set of Pareto-optimal operating policies in a single optimization run.
Numerical and practical advantages of MOFQI are demonstrated by application to both synthetic and real-world case studies.