Introduction to the Special Issue on Learning, Optimization, and Decision Making in DEDS

XI-REN CAO  
Hong Kong University of Science and Technology

The area of learning, optimization, and decision making in stochastic DEDS has witnessed a rapid development in recent years, primarily because of the availability of massive computation power and the demand in designing complex modern systems such as communication networks, manufacturing systems, and intelligent robots, etc. Researchers in many different disciplines, including control systems, computer sciences, operations research, artificial intelligence, and neural and cognitive sciences, etc, have made significant contributions. These research works share the same goal but are diverse in their perspectives, approaches, as well as concepts and terminology. For example, perturbation analysis (PA) of DEDS studies the optimization problem from a sensitivity view, and it provides estimates for performance sensitivities based on a single sample path of a DEDS; recent works on Markov decision problems (MDP) emphasize on approximating the value functions (e.g., neuro-dynamic programming); and in reinforcement learning (RL), decisions are made to improve a system’s performance by analyzing its behavior.

It has been realized that these different approaches are closely related. The goals of this special issue are to bring these research methodologies together in a somewhat unified way to the research community and to encourage cross-disciplinary cooperation. This is in the same spirit as the NSF sponsored workshop “Learning and Approximate Dynamic Programming” (Playacer, Mexico, April 2002), which brought people in different areas together “to learn new things from new people and communities we do not already have regular contact with, . . ., and to build new partnerships and such.” We trust that by taking an overall view of this fascinating subject, new insights, new research directions and solutions will start to emerge.

The paper “From Perturbation Analysis to Markov Decision Processes and Reinforcement Learning” by Xi-Ren Cao provides a sensitivity point of view of PA, MDP, and RL, and this new unified view on the relations among these areas brings in some new research directions. The paper “Recent Advances in Hierarchical Reinforcement Learning” by Barto and Mahadevan addresses the “curse of the dimensionality” issue by using an hierarchical learning approach; it is observed that this approach is related to hierarchical control of hybrid systems and the “time-aggregation” approach proposed in DEDS community. In “Policy Evaluation Algorithms with Linear Function Approximation,” Nedic and Bertsekas proposed two algorithms ($\lambda$-LSPE and LSTD ($\lambda$)) for approximating the cost-to-go function with discounted cost by using temporal difference with a linear approximation architecture. In “Approximate Gradient Methods in Policy-Space Optimization of Markov Reward Processes,” Marbach and Tsitsiklis briefly reviewed the stochastic gradient descent methods applied to optimization of parameterized MDPs and resulting algorithms, and addressed issue of variance reduction.
This is in the same spirit as the PA-based optimization. The paper "Simulation for Performance Evaluation and Policy Selection in Multiclass Networks" by Henderson, Meyn, and Tadic develops approximate modeling and policy synthesis methods for large multiclass networks by using the fluid model; the emphasis is also on the approximation of value functions and variance reduction. All these papers are written in a self-contained style so that the readers can get an overview as well as understand the current trend of these research areas.

In this special issue, we do not try to unify notations and terminology for two reasons. Firstly, it is not possible to do so at present because a one-to-one correspondence between the concepts in these different areas has not been completely established yet. Secondly and more importantly, different terminology represents different perspectives on the same concept; it is these different perspectives that motivate different thinking and hence are what precisely we want to keep. For example, the solution to the Poisson equation is called the potentials in PA, which represent the potential contribution of the current state to the performance and/or performance sensitivities; it is called a bias or different reward in MDP, which emphasizes the difference of the cost value with the steady state value; and it is also called a value function or cost-to-go function. Other examples are: the single sample path based approach in system theory is also called an on-line approach, which is equivalent to "learning" in RL; Poisson equation is the same as Bellman equation, etc.

We hope that this special issue may help to promote this fascinating research area and to facilitate cooperation among researchers in different disciplines.