Mach Learn (2011) 84:1–6
DOI 10.1007/s10994-011-5255-6

EDITORIAL

Introduction to the special issue on empirical evaluations in reinforcement learning

Shimon Whiteson · Michael L. Littman

Received: 13 May 2011 / Accepted: 19 May 2011 / Published online: 3 June 2011

The field of *reinforcement learning* aims to develop algorithms that use experience to optimize behavior in sequential decision problems such as (partially observable) Markov decision processes ((PO)MDPs). In such problems, an autonomous agent interacts with an external environment by selecting actions and seeks the sequence of actions that maximizes its long-term performance. In reinforcement learning, the environment is typically initially unknown and learning takes place *online*, that is, the agent's performance is assessed throughout learning instead of only afterwards.

Since many challenging and realistic tasks are well described as sequential decision problems, the development of effective reinforcement-learning algorithms plays an important role in artificial intelligence research. Recent years have seen enormous progress, both in the development of new methods and the theoretical understanding of existing methods. In particular, great strides have been made in approximating value functions, exploring efficiently, learning in the presence of multiple agents, coping with partial observability, inducing models, and reasoning hierarchically.

The focus of this special issue is not the development of new algorithms but the empirical evaluation of existing ones. Like other machine-learning methods, reinforcement-learning approaches are typically evaluated in one or more of the following three ways: (1) subjectively, (2) theoretically, and (3) empirically. Subjective evaluations, in which researchers assess the significance and potential of new ideas, are important because they leverage the powerful intuition of experts to guide the research process. However, they are also limited because they cannot validate ideas that go against such intuition; for example, they cannot expose fallacious assumptions. Theoretical evaluations are also important, as they are

S. Whiteson (✉)
Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
e-mail: s.a.whiteson@uva.nl

M.L. Littman
Department of Computer Science, Rutgers University, New Brunswick, NJ, USA
e-mail: mlittman@cs.rutgers.edu

perhaps the most rigorous way to assess an algorithm. However, for many problems of real-world interest, obtaining theoretical results is extremely difficult. In addition, even when obtained, such results may not accurately predict performance in practice, because either bounds are loose or assumptions do not strictly hold.

Thus, empirical evaluations play an important role. Just as competitions such as the Netflix Prize (Bell et al. 2010) and the DARPA Grand Challenge (Montemerlo et al. 2006; Seetharaman et al. 2006) can spur rapid progress, careful empirical evaluations can stimulate the development of increasingly practical algorithms (Cohen and Howe 1988; Langley 1988; Simon 1995). For example, the UCI repository (Asuncion and Newman 2007) of benchmark learning problems has helped supervised learning methods transition from a laboratory curiosity to a critical component in real-world applications.

However, empirically evaluating the performance of a reinforcement-learning algorithm poses unique challenges. There are many ways to define the problem and evaluate performance. In addition, a benchmark task consists not of a fixed data set that can be easily published, but of a dynamic module, such as a simulator, with which different candidate agents can interact. Consequently, the task of sharing benchmarks is difficult, as completely describing such modules in an article is often infeasible (White 2006).

Fortunately, in recent years, various researchers have taken initial steps towards addressing these issues. Perhaps most significant from a practical perspective is the development of RL-Glue (Tanner and White 2009), a task-, language-, and platform-independent protocol for connecting agents and environments. In addition, the recent international Reinforcement Learning Competitions (Whiteson et al. 2010), which used RL-Glue as an evaluation platform, have helped focus attention on specific benchmark problems and assess the current state of the art. Finally, researchers have begun to address the 'meta' question of how to find good methodologies for evaluating reinforcement-learning methods (Nouri et al. 2009; Whiteson et al. 2009, 2011).

Our primary aim in editing this special issue is to bring together and highlight recent research developments that address important questions relevant to empirical evaluations in reinforcement learning. In the remainder of this editorial, we enumerate several of these questions and describe how the articles in the special issue address them.

1 How should software for empirical research be designed?

Doing careful experiments in reinforcement learning requires designing suitable software. While RL-Glue has proved a useful software protocol, it is far from a panacea. In their article, "On the Analysis and Design of Software for Reinforcement Learning, with a Survey of Existing Systems", Kovacs and Egginton investigate practical issues involved in constructing software for reinforcement-learning experiments. They argue that libraries of reusable software components are of greater value than repositories of ready-made agents and environments, since they facilitate the testing of multiple learning algorithms under multiple settings. In addition, they present "Alfred's Story", an illustration of the software-engineering challenges a typical reinforcement-learning researcher faces when building an experimental setup from scratch and then expanding it to handle different scenarios and interact with components written by others. They analyze the requirements of a reinforcement-learning system and argue that the process of designing such a system can yield insights into the scope of reinforcement learning as well as the relationships between the algorithms and problems within that scope. Finally, they present a comprehensive survey of existing systems along with recommendations about which systems are best suited for various users.

2 What are the right evaluation methodologies?

Empirically evaluating reinforcement-learning algorithms requires a suitable methodology for conducting trials and quantifying performance. The resulting challenges are addressed by several articles.

In “Empirical Evaluation Methods for Multiobjective Reinforcement Learning Algorithms”, Vamplew, Dazeley, Berry, Issabekov, and Dekker consider reinforcement-learning problems that require the simultaneous satisfaction of multiple reward functions. Empirically evaluating performance in such problems is challenging in part because multi-objective reinforcement-learning methods differ widely with respect to the problem settings they assume and the goals they aim to reach. Vamplew et al. argue that different approaches to evaluation are needed when evaluating *multiple-policy* methods, which produce a set of all Pareto-optimal polices or an approximation thereof, than when evaluating *single-policy* methods, which produce one policy that maximizes a linear scalarization of the objectives or satisfies constraints on those objectives. For the former, they point out limitations in the simple scalar metrics and graphical representations in common use and propose instead the use of learning curves that track the hypervolume of the space dominated by the Pareto front. For the latter, they sketch several options, including hypervolume metrics and multi-objective regret, as well as metrics based on experiments with real or simulated users.

Fern, Khardon, and Tadepalli also address the issue of evaluation methodologies in “The First Learning Track of the International Planning Competition”. They consider a problem closely related to traditional reinforcement learning; specifically, they examine making high-quality decisions in an initially unknown environment given a formal description of its dynamics. They describe a recent competition in which closely related planning problems were presented to a planning system that can change its behavior on future problems on the basis of its experience. They describe the evaluation methodologies they developed to ensure rigorous evaluation in this unique setting. In particular, they emphasize the importance of selecting a performance metric that truly aligns with the field’s goals. Otherwise, the algorithms that we consider “good” may end up losing to algorithms that are overfit to the metric.

In their article, “Informing Sequential Clinical Decision-Making through Reinforcement Learning: An Empirical Study”, Shortreed, Laber, Lizotte, Stroup, Pineau, and Murphy consider the design of treatment plans for managing chronic diseases. As their task is to find an optimal sequence of decisions given data, it is a natural fit for reinforcement-learning techniques. However, the special constraints of this application domain make it particularly challenging. While the treatment history can be viewed as the state and possible interventions are actions, it is not clear how best to formulate reward. The choice can have a direct impact on the quality of life of patients who adopt the automatically generated plans. The authors also consider how a physician or patient can be confident that the computed treatment policy is well justified. They examine techniques for computing confidence intervals on the learned value function, which can be used to increase confidence in the individual decisions that are recommended.

3 What are suitable benchmark tasks?

Even given appropriate software and evaluation methodologies, researchers doing empirical comparisons must still select suitable benchmark tasks. This issue is addressed by several articles in the special issue. Vamplew et al. propose several benchmark tasks for the multi-objective setting. Fern et al. describe the benchmark tasks used in the learning track of the planning competition.

The issue is given thorough treatment by Hafner and Riedmiller in "Reinforcement Learning in Feedback Control: Challenges and Benchmarks from Technical Process Control". The authors propose four benchmark reinforcement-learning problems drawn from the area of technical process control. These tasks, which they call Underwater Vehicle, Pitch Control, Magnetic Levitation, and Heating Coil, can be used to compare human-designed control rules with those learned from experience. The domains were chosen to highlight several core issues that are important in process control environments: the necessity of going beyond simple linear control rules; the prevalence of long-range dynamic effects; the requirement for highly precise control laws; the ubiquity of external variables; and the criticality of dealing with changing setpoints. The authors propose two measures for controller quality: how well the learned controller performs and how much experience it needs to reach a target performance level.

Kober and Peters also explore the issue of how to select suitable benchmarks in their article, "Policy Search for Motor Primitives in Robotics". In evaluating policy-search methods for motor-control tasks, they underscore the limitations of synthetic benchmarks, which often unrealistically assume discrete states and actions and the availability of millions of samples for learning. While motor-control tasks can serve as more realistic benchmarks, conducting empirical evaluations in such settings is not easy. Experiments on real robots are difficult to repeat, since they typically depend on elements that are hard to share, such as custom hardware and human demonstrations used to initialize learning. Testing algorithms in simulation first can help, but making simulators accurate enough is challenging. The authors suggest that *energy-absorbing* scenarios, where the robot absorbs energy from its actions, are typically so difficult to simulate that it is easier to learn on a real robot than to learn first in simulation and transfer the results. In contrast, in *energy-emitting* scenarios such transfer is often feasible, especially if noise is added to the simulator. In borderline scenarios, the learning parameters optimized in simulation may transfer even when the learned policies themselves do not.

4 How do current methods fare in empirical comparisons?

In addition to discussing how empirical evaluations should be done, most of the articles in this special issue also present the results of such comparisons. Vamplew et al. present the results of comparing several standard multi-objective methods on their proposed benchmarks using their proposed evaluation methodologies. Similarly, Hafner and Riedmiller evaluate the performance of their own method, Neural Fitted Q Iteration with Continuous Actions, on their proposed benchmark tasks and encourage the reader to apply his or her favorite approach to see how it compares. Shortreed et al. also describe an empirical study of reinforcement-learning approaches applied to actual clinical trial data carried out by some of the trailblazers of the emerging field of sequential clinical decision making.

Fern et al. describe in detail the results of the learning track of the planning competition. They found that, like the best Netflix Prize systems, ensemble methods that combine the strengths of a diverse set of powerful planners can be tenacious competitors. While several learning systems were able to improve with exposure to related problems within a domain, the final performance achieved by learning systems was on par with the best non-learning planners. That is, one of the best ways to solve a series of related planning problems is still to treat each of them independently, without learning. There is reason to hope, however, that the evaluation mechanisms they propose will help spur the field to breakthroughs in the near term.

Kober and Peters present extensive empirical results evaluating the performance of a range of policy-search methods on a series of simulated and real motor-control tasks. They propose a general framework for policy-search methods in which policy improvement is formulated as a maximization of a lower bound on the expected return. They show that many existing episodic methods, including REINFORCE, policy gradient methods, and Natural Actor-Critic, follow from this framework. They also derive a new method called Policy Learning by Weighting Exploration with the Returns (PoWER), which relies on expectation-maximization instead of gradients and performs structured, state-dependent exploration. By coupling these methods with basis functions derived from motor primitives, they apply them to several motor-control tasks such as Underactuated Swing-Up and Ball-in-a-Cup in simulation and, in some cases, on real robots. The results they present both verify the efficacy of the PoWER method and illustrate the state of the art in reinforcement learning for motor control.

In “Characterizing Reinforcement Learning Methods through Parameterized Learning Problems”, Kalyanakrishnan and Stone develop a parameterized class of reinforcement-learning problems and use it to conduct an extensive empirical analysis of the strengths and weaknesses of value-function and policy-search methods. Motivated by the lack of a “field guide” for reinforcement learning, they offer results that can inform the development of “rules of thumb” for practitioners.

Their framework allows controlled tests of the effects of five critical factors: state space size, transition stochasticity, partial observability, and the expressiveness and generalization of function approximation. They compare multiple value-function and policy-search methods and present detailed results on Sarsa and CMA-ES, the best performing in each category. While these results offer additional confirmation for previously observed trends, for example, that Sarsa tends to learn more quickly, they are also qualitatively novel. Unlike previous empirical studies, their framework allows the same function approximation representations to be used by both types of methods, enabling a more controlled examination of the effect of such representations on relative performance. This setup leads to new insights, for example, that CMA-ES can be much more robust when the representation is highly deficient.

5 Conclusion

We believe that the question of how best to conduct empirical evaluations in reinforcement learning remains a critical research topic. The research described in this special issue shows that promising initial progress has been made towards the development of principled answers. The results of the empirical evaluations presented by these articles not only offer new insights about the strengths and weaknesses of current reinforcement-learning methods, but also underscore the capacity of careful evaluations to generate such insights. We hope that this special issue will serve as a foundation for and stimulus to further developments in reinforcement-learning software, evaluation methodologies, benchmark tasks, and empirical comparisons.

References

Asuncion, A., & Newman, D. (2007). UCI machine learning repository. http://www.ics.uci.edu/mlearn/MLRepository.html.

Bell, R. M., Koren, Y., & Volinsky, C. (2010). All together now: a perspective on the Netflix Prize. *Chance*, *23*(1), 24–29.

Cohen, P., & Howe, A. (1988). How evaluation guides AI research. *AI Magazine*, *9*(4), 35–43.

Langley, P. (1988). Machine learning as an experimental science. *Machine Learning*, *3*(1), 5–8.

Montemerlo, M., Thrun, S., Dahlkamp, H., Stavens, D., & Strohband, S. (2006). Winning the DARPA Grand Challenge with an AI robot. In *Proceedings of the national conference on artificial intelligence* (pp. 982–987).

Nouri, A., Littman, M. L., Li, L., Parr, R., Painter-Wakefield, C., & Taylor, G. (2009). A novel benchmark methodology and data repository for real-life reinforcement learning. In *Proceedings of the 26th international conference on machine learning*. Poster.

Seetharaman, G., Lakhotia, A., & Blasch, E. (2006). Unmanned vehicles come of age: the DARPA grand challenge. *Computer*, *39*(12), 26–29.

Simon, H. (1995). Artificial intelligence: an empirical science. *Artificial Intelligence*, *77*(1), 95–127.

Tanner, B., & White, A. (2009). RL-Glue: Language-independent software for reinforcement-learning experiments. *Journal of Machine Learning Research*, *10*, 2133–2136.

White, A. (2006). *A standard system for benchmarking in reinforcement learning*. Master's thesis, University of Alberta, Alberta, Canada.

Whiteson, S., Tanner, B., Taylor, M. E., & Stone, P. (2009). Generalized domains for empirical evaluations in reinforcement learning. In *ICML 2009: proceedings of the twenty-sixth international conference on machine learning: workshop on evaluation methods for machine learning*.

Whiteson, S., Tanner, B., Taylor, M. E., & Stone, P. (2011). Protecting against evaluation overfitting in empirical reinforcement learning. In *ADPRL 2011: proceedings of the IEEE symposium on adaptive dynamic programming and reinforcement learning* (pp. 120–127).

Whiteson, S., Tanner, B., & White, A. (2010). The reinforcement learning competitions. *AI Magazine*, *31*(2), 81–94.