The Reinforcement Learning Competition 2014

Christos Dimitrakakis, Guangliang Li, Nikolaos Tziortziotis

■ *Reinforcement learning is one of the most general problems in artificial intelligence. It has been used to model problems in automated experiment design, control, economics, game playing, scheduling, and telecommunications. The aim of the reinforcement learning competition is to encourage the development of very general learning agents for arbitrary reinforcement learning problems and to provide a test bed for the unbiased evaluation of algorithms.*

Reinforcement learning (RL) is the problem of learning how to act only from interaction and limited reinforcement. An agent takes actions in an unknown environment, observes their effects, and obtains rewards. The agent's aim is to learn how the environment works in order to maximize the total reward obtained during its lifetime. RL problems are quite general. They can subsume almost any artificial intelligence problem, through suitable choices of environment and reward. For example, learning to play games can be formalized as a problem where the environment is adversarial and there is a positive (negative) reward when the agent wins (loses). An introduction to RL is given by the book of Sutton and Barto (1998), while a good overview of recent algorithms is given by Szepesvári (2010).

Most RL problems are formalized by modeling the interaction between the agent and the environment as a discrete-time process. At each time t, the agent takes an action at from some action set A and obtains an observation $x_{t+1} \in X$ and a scalar reward $r_{t+1} \in \mathbb{R}$. Part of the problem is estimating a model for the environment, that is, learning how the observations and rewards depend upon the actions taken. However, the agent's actual objective is to maximize the total reward

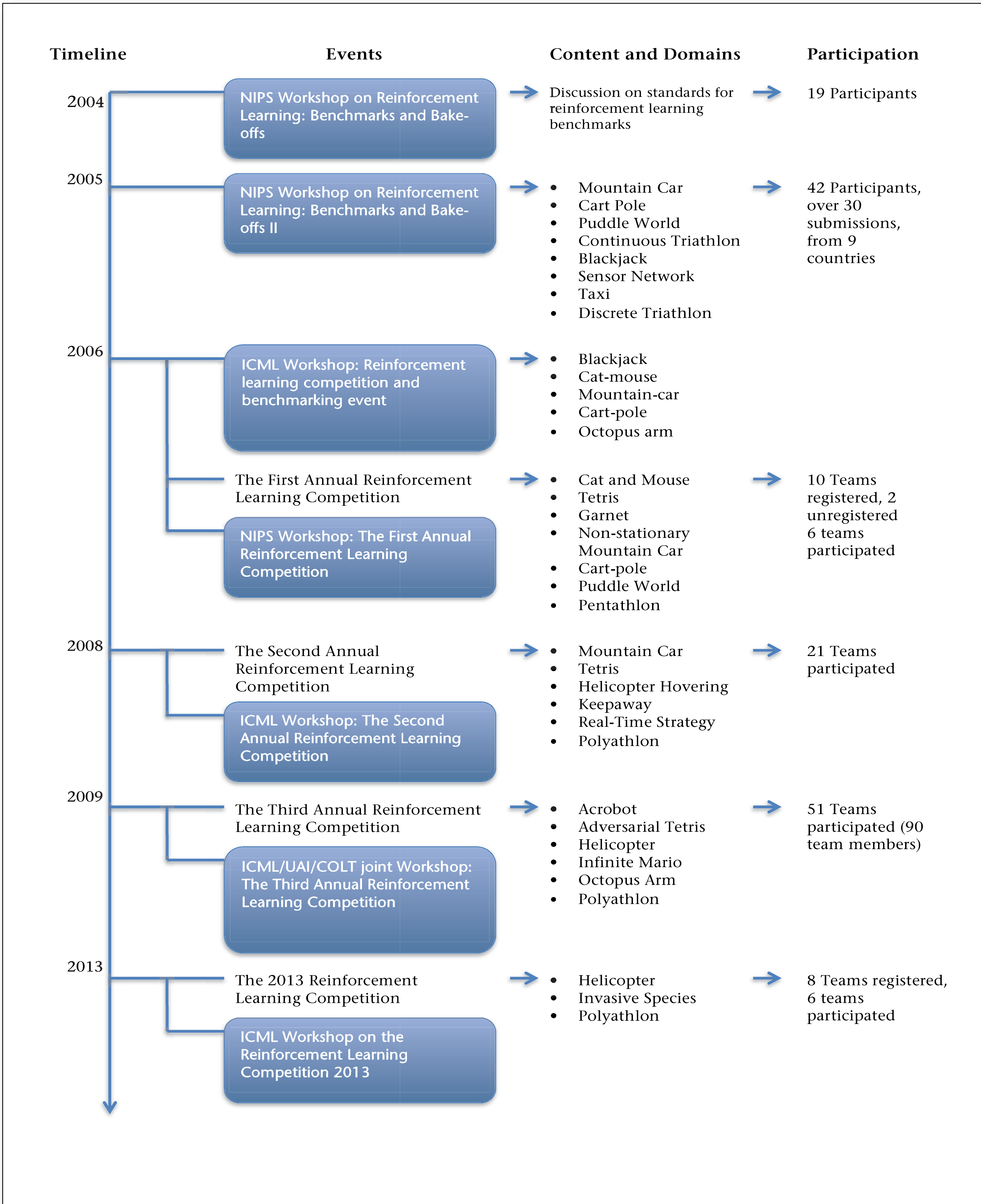

Figure 1. History of Reinforcement Learning Competition and Collocated Events.

$$U = \sum_{t=1}^{T} r_t$$

during its lifetime. Thus, it must not waste a lot of time learning aboutenvironment details that are not relevant for obtaining rewards. It should properly balance exploitation of what is already known with exploration of unknown parts of the environment. This exploration-exploitation dilemma is at the core of the reinforcement learning problem, and it is due to its interactive nature.

Comparisons between different RL algorithms are inherently difficult. Authors of algorithms may spend a lot of time tuning them to work well in specific types of domains, even when they are ostensibly learning from scratch. What is needed is a way to perform unbiased comparisons. This is offered by the RL competition. While unbiased comparisons for static data sets are a well-established practice, the interactive nature of RL means that no fixed data sets are possible. However, the same principle applies to complete environments. A designer can tune algorithms on a small set of representative environments, and then use another set of similar, but nonidentical environments to test them on.

This is the principle used by the reinforcement learning competition, through the use of generalized domains. Each domain corresponds to a different type of problem, such as helicopter control. Training and testing is performed on specific domain instances of the problem, for example, helicopters with different physical characteristics. The competition format ensures an unbiased comparison between algorithms, that truly tests their ability to learn. The next section gives an overview of the competition's history and its evolution to the current format. The goals and format of the upcoming competition are detailed in the 2014 competition section.

History of the Competition

The competition aims to foster development of general RL algorithms and to promote their systematic comparison. It has generated a set of benchmark domains and results, which were used as a basis of comparisons in further work. The competition and the related open source code have been used in teaching and research projects (Whiteson, Tanner, and White 2010) and have greatly increased the interest and focus in RL by clarifying its objectives and challenges, making the area more exciting and enjoyable, especially for novices in the RL community.

The competition history is summarized in figure 1, starting with a workshop at the 2004 Neural Information Processing Systems (NIPS) conference, organized by Rich Sutton and Michael Littman. This event discussed a standard set of benchmarks and a series of competitive events to enhance reinforcement learning research. The next event was held at NIPS 2005, with more than 30 teams participating. This event featured a new RL framework (later called RL-Glue) developed at the University of Alberta. This offers a language independent API for evaluating agents in different environments and experimental setups (Tanner and White 2009). RL-Glue was used by all later competitions and events.

After another event at the International Conference on Machine Learning (ICML) in 2006, organized by Shie Mannor and Doina Precup, the first annual RL competition was held at NIPS 2006. This was the first official competition with prizes and winner certificates. Eight teams participated, and the winners presented their work at the workshop.

The second annual RL competition was held by Shimon Whiteson, Brian Tanner, and Adam White in 2008 at an ICML workshop, in which 21 teams participated (Whiteson, Tanner, and White 2010). After this success, the third reinforcement learning competition was held in 2009 with a joint workshop at ICML, Uncertainty in Artificial Intelligence (UAI), and Conference on Learning Theory (COLT). The competition attracted 51 teams comprising 90 members. Prizes included electronic devices and gift certificates, while travel scholarships were available for participants to defray the cost of attendance.

The 2008 and 2009 competition had higher levels of participation than previous events. However, the competition was not revived until 2013, with an associated ICML workshop, to ensure its continuing existence.

Domains

The competition includes domains with various characteristics and complexity to encourage researchers to develop methods for challenging problems. The 2005 event featured three continuous domains (Mountain Car, Puddle World, and Cart Pole) and three discrete domains (Blackjack, Sensor Network, and Taxi). In addition, there were two triathlon events, in which the same agent competed on all the three continuous or discrete domains.

The 2006 competition added the adversarial cat-mouse domain and octopus arm control. The latter is a large-scale domain with high-dimensional continuous observation and action spaces (Yekutieli et al. 2005). However, the proliferation of domains meant that many domains had very few entrants. For that reason, the total number was reduced steadily, but new domains were introduced every year. The 2008 competition featured Helicopter Hovering (Bagnell and Schneider 2001; Ng et al. 2006), Keepaway Soccer (Stone et al. 2006; Stone, Sutton, and Kuhlmann 2005), and Real-Time Strategy. Helicopter Hovering and Keepaway were high-dimensional domains, which were problematic for traditional RL methods. The 2009 competition introduced the classic Acrobot problem (DeJong and Spong 1994) and the Infinite

Mario Game. Finally, the 2013 competition introduced the Invasive Species Problem (Muneepeerakul et al. 2007), a very high-dimensional structured domain.

Polyathlon domains are a special characteristic of the RL competition, starting with the triathlon events of 2006. In those, agents are tested in a series of tasks, which could be completely unrelated and are not known in advance. Consequently, entrants must submit very general agents. They must be able to learn to act optimally in essentially arbitrary environments, while exploiting whatever structure they can discover.

Evaluation of Agent Performance

In the competition, the agent performance is based on the total reward the agent received. At the first benchmarking event in 2005, there was no separation between training and testing, and consequently allowed considerable prior knowledge to be used (Whiteson, Tanner, and White 2010).

To promote agents that really learn, the 2006 event separated agent evaluation into training and testing phases. During training, participants were allowed unlimited runs on a fixed training set of problems. The testing phase used a different but similar set of problems. Agents were allowed a period of free exploration in the test set, which did not count toward the final score. Later events separated this out in an explicit proving phase, where entrants had a limited number of attempts to try out their agents in new environments. The combination of this with a leaderboard, which compared the results of different teams, motivated participants to continually improve their agents throughout the competition.

The 2014 Competition

The organizers of the 2014 competition are Christos Dimitrakakis (Chalmers University of Technology), Guangliang Li (University of Amsterdam), and Nikolaos Tziortziotis (University of Ioannina). The competition server will open this fall, and the competition will take place in November–December 2014. Competitors will have the opportunity to compete in domains with various characteristics (such as complexity, difficulty, and others) in an unbiased manner. At the end of the competition, authors will be requested to submit a paper to the competition workshop, collocated with a major conference in the AI field, present their methods, and receive their prizes.

Competition Format

The competition consists of three main phases: training, proving, and testing. During the first two phases the participants will have the opportunity to refine their agents. Only the results of the testing phase will be taken into account for evaluation. As in the last four competitions, our evaluation criterion will be the online performance of the agents: the total reward collected. As the testing instances will differ from both the training and proving instances, this format benefits agents that can learn efficiently online.

In the training phase, competitors receive a set of training environments, which are specific instances of the generalized domain. For the helicopter example, this would mean that the physical characteristics of the helicopter may vary across instances. The competitors can use the training environments to make their algorithms as robust as possible. This local development allows an unlimited number of attempts.

In the proving phase of the competition, competitors can try their algorithms in preliminary experiments in new instances on the competition server. They can thus verify that their algorithms can generalize well outside the training set. The performance of various competitors in this phase is visible in a leaderboard, and helps to ensure that competitors can perform well. The number of attempts per week in this phase is limited. However, the scores during this phase do not count toward the final competition score.

The final competition phase is testing. During this phase, competitors test their algorithms in new domain instances on the competition server. The score during this phase is used for the agent's evaluation. The scores across different domains are treated independently, and every team is allowed to compete in multiple domains.

The competition platform is based on the RL-Glue[1] API. This is a standard interface for connecting RL agents and environments and allows participants to create their agents in their language of choice. Apart from the training phase, participants connect to the central competition server over the Internet, to obtain new environments and report results. During proving and testing, the competition software updates the leaderboards on the web. In this way, participants can see how their agents compare with others.

Getting Involved.

We are actively looking for volunteers to help with various aspects of the competition. First, we welcome the donation of prizes to foster additional involvement in the competition. These can include travel scholarships, monetary prizes, or trophies. Second, we would like to invite volunteers to help with technical issues. To ensure the continuity of the competition, we request enthusiastic individuals that can help with the technical side of the competition for 2014. Third, we request submissions of new domains. You are welcome to email us your suggestions. Of particular interest for the competition this year will be (1) adversarial problems and (2) games problems with changing tasks. Finally, and most importantly, we encourage your participation with agent submissions. We hope to have a vibrant and exciting 2014 competition.

References

Bagnell, J. A., and Schneider, J. G. 2001. Autonomous Helicopter Control Using Reinforcement Learning Policy Search Methods. In *Proceedings of the IEEE International Conference on Robotics and Automation,* volume 2, 1615–1620. Piscataway, NJ: Institute for Electrical and Electronics Engineers.

DeJong, G., and Spong, M. W. 1994. Swinging Up the Acrobot: An Example of Intelligent Control. In *Proceedings of the American Control Conference*, volume 2, 2158–2162. Piscataway, NJ: Institute for Electrical and Electronics Engineers.

Muneepeerakul, R.; Weitz, J. S.; Levin, S. A.; Rinaldo, A.; and Rodriguez-Iturbe, I. 2007. A Neutral Metapopulation Model of Biodiversity in River Networks. *Journal of Theoretical Biology* 245(2): 351–363. dx.doi.org/10.1016/j.jtbi.2006.10.005

Ng, A. Y.; Coates, A.; Diel, M.; Ganapathi, V.; Schulte, J.; Tse, B.; Berger, E.; and Liang, E. 2006. Autonomous Inverted Helicopter Flight Via Reinforcement Learning. In *Experimental Robotics IX,* 363–372. Berlin: Springer.

Stone, P.; Kuhlmann, G.; Taylor, M. E.; and Liu, Y. 2006. Keepaway Soccer: From Machine Learning Testbed to Benchmark, 93–105. In *RoboCup 2005: Robot Soccer World Cup IX*. Berlin: Springer.

Stone, P.; Sutton, R. S.; and Kuhlmann, G. 2005. Reinforcement Learning for Robocup Soccer Keepaway. *Adaptive Behavior* 13(3): 165–188. dx.doi.org/10.1016/j.jtbi.2006.10.005

Sutton, R. S., and Barto, A. G. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press.

Szepesvári, C. 2010. Algorithms for Reinforcement Learning. In *Synthesis Lectures on Artificial Intelligence and Machine Learning,* volume 4. San Rafael, CA: Morgan & Claypool Publishers.

Tanner, B., and White, A. 2009. RL-Glue: Language-Independent Software for Reinforcement-Learning Experiments. *The Journal of Machine Learning Research* 10: 2133–2136.

Whiteson, S.; Tanner, B.; and White, A. 2010. Report on the 2008 Reinforcement Learning Competition. *AI Magazine* 31(2): 81.

Yekutieli, Y.; Sagiv-Zohar, R.; Aharonov, R.; Engel, Y.; Hochner, B.; and Flash, T. 2005. Dynamic Model of the Octopus Arm. I. Biomechanics of the Octopus Reaching Movement. *Journal of Neurophysiology* 94(2): 1443–1458. dx.doi.org/10.1152/jn.00684.2004

Christos Dimitrakakis is a researcher and lecturer at Chalmers University of Technology. His research interests include decision theory and reinforcement learning, security, and privacy. He obtained his Ph.D. in 2006 from EPFL, Switzerland.

Guangliang Li is a Ph.D. candidate at the University of Amsterdam. His research focuses on reinforcement learning and human computer interaction.

Nikolaos Tziortziotis is a Ph.D. candidate at University of Ioannina. His research interests include machine learning and reinforcement learning.