Progress Report Towards PhD Degree, Year Three

Research Topic: Rational Metareasoning in Problem Solving Search

David Tolpin

Department of Computer Science

Ben-Gurion University of the Negev

Under the supervision of Professor Solomon Eyal Shimony

August 4, 2013

1 Research Progress

During the third year of my Ph.D. studies I worked on Monte Carlo Tree Search, in particular on VOI-aware MCTS for simple regret.

MCTS, and especially UCT [5] appears in numerous search applications, such as [3]. Although these methods are shown to be successful empirically, most authors appear to be using UCT "because it has been shown to be successful in the past", and "because it does a good job of trading off exploration and exploitation". While the latter statement may be correct for the Multi-armed Bandit problem and for the UCB1 algorithm [1], we argue that a simple reconsideration from basic principles can result in schemes that outperform UCT.

The core issue is that in MCTS for adversarial search and search in "games against nature" the goal is typically to find the best first action of a good (or even optimal) policy, which is

closer to minimizing the simple regret, rather than the cumulative regret minimized by UCB1. However, the simple and the cumulative regret cannot be minimized simultaneously; moreover, [2] shows that in many cases the smaller the cumulative regret, the greater the simple regret.

In a series of papers [6, 4, 7], we proposed a new MCTS sampling scheme, SR+CR, which is based on optimizing the simple regret of sampling, and outperforms UCT empirically. Further on, we derived theoretical distribution-independent bounds for the blinkered VOI estimate, and improved the sampling scheme through the use of the bound-based VOI estimate. The new sampling scheme was emprically evaluated on a number of domains, including POMDP and Computer Go, and outperformed other known sampling schemes.

The research was presented at the leading conferences in the field of Artificial Intelligence: AAAI-2012, UAI-2012, ECAI-2012.

2 Publications

References

[1] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the Multiarmed bandit problem. *Mach. Learn.*, 47:235–256, May 2002.

[2] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theor. Comput. Sci.*, 412(19):1832–1852, 2011.

[3] Patrick Eyerich, Thomas Keller, and Malte Helmert. High-quality policies for the canadian travelers problem. In *In Proc. AAAI 2010*, pages 51–58, 2010.

[4] Nicholas Hay, Stuart J. Russell, David Tolpin, and Solomon Eyal Shimony. Selecting computations: Theory and applications. In *UAI*, pages 346–355, 2012.

[5] Levente Kocsis and Csaba Szepesvári. Bandit based Monte-Carlo planning. In *ECML*, pages 282–293, 2006.

[6] David Tolpin and Solomon Eyal Shimony. MCTS based on simple regret. In *AAAI*, pages 570–576. AAAI Press, 2012.

[7] David Tolpin and Solomon Eyal Shimony. VOI-aware MCTS. In *ECAI*, pages 929–930, 2012.